```
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 7
BQ248646/c
LOCUS       BQ248646                566 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25006C10F TaE25 Triticum aestivum cDNA clone TaE25006C10F, mRNA
            sequence.
ACCESSION   BQ248646
VERSION     BQ248646.1  GI:20444522
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 006  row: C  column: 10
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .566
                      /organism="Triticum aestivum"
                      /cultivar="Glenlea"
                      /db_xref="taxon:4565"
                      /clone="TaE25006C10F"
                      /clone_lib="TaE25"
                      /tissue_type="developing seeds"
                      /dev_stage="25 days after anthesis"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                      Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                      of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      136 a    138 c    143 g    149 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 566;
  Best Local Similarity   93.2%;  Pred. No. 5.2e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 8
BQ248555/c
LOCUS       BQ248555                575 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25007E01F TaE25 Triticum aestivum cDNA clone TaE25007E01F, mRNA
            sequence.
ACCESSION   BQ248555
VERSION     BQ248555.1  GI:20444431
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 007  row: E  column: 01
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .575
                      /organism="Triticum aestivum"
                      /cultivar="Glenlea"
                      /db_xref="taxon:4565"
                      /clone="TaE25007E01F"
                      /clone_lib="TaE25"
                      /tissue_type="developing seeds"
                      /dev_stage="25 days after anthesis"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                      Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                      of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      141 a    140 c    142 g    151 t      1 others
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 575;
  Best Local Similarity   93.2%;  Pred. No. 5.3e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db    230 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 172


RESULT 9
BQ248056/c
LOCUS       BQ248056                593 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25013F06F TaE25 Triticum aestivum cDNA clone TaE25013F06F, mRNA
            sequence.
ACCESSION   BQ248056
VERSION     BQ248056.1  GI:20443932
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 013  row: F  column: 06
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .593
                      /organism="Triticum aestivum"
                      /cultivar="Glenlea"
                      /db_xref="taxon:4565"
```

```
                     /clone="TaE25013F06F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      145 a    146 c    147 g    155 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 593;
  Best Local Similarity   93.2%;  Pred. No. 5.3e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 10
B68191/c
LOCUS       B68191                   610 bp    DNA     linear   GSS 18-JUN-1998
DEFINITION  CIT978SK-A-492C12.TV CIT978SK Homo sapiens genomic clone A-492C12,
            DNA sequence.
ACCESSION   B68191
VERSION     B68191.1  GI:2666901
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Kim,U.-J., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a human BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT978SK-A-492C12.TP
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Homo sapiens"
                     /db_xref="GDB:726476"
                     /db_xref="taxon:9606"
                     /clone="A-492C12"
                     /clone_lib="CIT978SK"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
                     CalTech Human BAC Library A"
BASE COUNT      146 a    159 c    165 g    140 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 17;  Length 610;
  Best Local Similarity   93.2%;  Pred. No. 5.4e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     288 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 230


RESULT 11
BQ248255/c
LOCUS       BQ248255                 625 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25011C08F TaE25 Triticum aestivum cDNA clone TaE25011C08F, mRNA
            sequence.
ACCESSION   BQ248255
VERSION     BQ248255.1  GI:20444131
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 011  row: C  column: 08
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25011C08F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      155 a    152 c    154 g    164 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 625;
  Best Local Similarity   93.2%;  Pred. No. 5.4e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 12
BQ248467/c
LOCUS       BQ248467                 625 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25008F09F TaE25 Triticum aestivum cDNA clone TaE25008F09F, mRNA
            sequence.
ACCESSION   BQ248467
VERSION     BQ248467.1  GI:20444343
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
```

```
           Cereal Research Centre, Agriculture and Agri-food Canada
           195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
           Tel: (204) 983-2340
           Fax: (204) 983-4604
           Email: scloutier@em.agr.ca
           was cloned directionally, not all sequences generated with reverse
           primer were from the 5' end (same with forward primer and 3' end).
           Average insert size is >870 bp
           Plate: 008  row: F  column: 09
           Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25008F09F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
  E COUNT      155 a    152 c    155 g    163 t
...IGIN

  Query Match           71.1%;  Score 52.6;  DB 14;  Length 625;
  Best Local Similarity  93.2%;  Pred. No. 5.4e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||| |||||||||||||||||||||||||||||| | || |
Db    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 13
BQ248453/c
LOCUS       BQ248453                630 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25008H04F TaE25 Triticum aestivum cDNA clone TaE25008H04F, mRNA
            sequence.
ACCESSION   BQ248453
VERSION     BQ248453.1  GI:20444329
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
  MENT      Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 008  row: H  column: 04
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25008H04F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      156 a    154 c    155 g    165 t
ORIGIN

  Query Match           71.1%;  Score 52.6;  DB 14;  Length 630;
  Best Local Similarity  93.2%;  Pred. No. 5.5e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||| |||||||||||||||||||||||||||||| | || |
Db    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 14
BQ248173/c
LOCUS       BQ248173                631 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25012C11F TaE25 Triticum aestivum cDNA clone TaE25012C11F, mRNA
            sequence.
ACCESSION   BQ248173
VERSION     BQ248173.1  GI:20444049
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 012  row: C  column: 11
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25012C11F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      156 a    154 c    155 g    166 t
ORIGIN

  Query Match           71.1%;  Score 52.6;  DB 14;  Length 631;
  Best Local Similarity  93.2%;  Pred. No. 5.5e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||| |||||||||||||||||||||||||||||| | || |
Db    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 15
BQ248939/c
LOCUS       BQ248939                633 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25002C05F TaE25 Triticum aestivum cDNA clone TaE25002C05F, mRNA
            sequence.
```

```
ACCESSION   BQ248939
VERSION     BQ248939.1  GI:20444815
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 002  row: C  column: 05
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25002C05F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      157 a    154 c    154 g    168 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 633;
  Best Local Similarity   93.2%;  Pred. No. 5.5e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||| |||||||||||||||||||||||||||||| | || |
 b    228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


Search completed: March  6, 2003, 00:22:23
Job time : 560.372 secs
```

```
VERSION     BQ248969.1  GI:20444845
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 001  row: G  column: 10
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25001G10F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT       88 a     99 c     96 g    105 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 388;
  Best Local Similarity   93.2%;  Pred. No. 4.5e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 5
BQ248026/c
LOCUS       BQ248026                502 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25014A07F TaE25 Triticum aestivum cDNA clone TaE25014A07F, mRNA
            sequence.
ACCESSION   BQ248026
VERSION     BQ248026.1  GI:20443902
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 014  row: A  column: 07
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25014A07F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      121 a    128 c    122 g    131 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 502;
  Best Local Similarity   93.2%;  Pred. No. 5e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     228 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 170


RESULT 6
BQ248278/c
LOCUS       BQ248278                535 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25011A03F TaE25 Triticum aestivum cDNA clone TaE25011A03F, mRNA
            sequence.
ACCESSION   BQ248278
VERSION     BQ248278.1  GI:20444154
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 011  row: A  column: 03
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25011A03F"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      130 a    133 c    132 g    139 t      1 others
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 14;  Length 535;
  Best Local Similarity   93.2%;  Pred. No. 5.1e-09;
```

```
          Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="RZPD clone CHBOp576F21214Q3"
                     /clone_lib="Zebrafish Kidney cDNA random primed, RZPD
                     library no: 576"
                     /dev_stage="adult"
                     /note="Organ: kidney; Vector: pBK-CMV"
BASE COUNT      163 a    165 c    162 g    165 t      2 others
ORIGIN

  Query Match             71.9%;  Score 53.2;  DB 9;  Length 657;
  Best Local Similarity   87.9%;  Pred. No. 3.3e-09;
  Matches   58;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTCA 61
           |||||||||||||||||| |||||||||||||||| ||||||||||||||| | |  |
Db     369 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCT 310

Qy      62 AACCGC 67
           | ||||
Db     309 AGCCGC 304


RESULT 2
BE661741/c
LOCUS       BE661741                 703 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  138t7 GmaxSC Glycine max cDNA, mRNA sequence.
ACCESSION   BE661741
VERSION     BE661741.1  GI:9987633
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Harris,N., Chapman,B.P. and Gijzen,M.
  TITLE     Gene expression in developing soybean seed coats
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Gijzen M
            Agriculture and Agri-Food Canada
            1391 Sandford Street, London, Ontario, Canada N5V 4T3
            Tel: 519 457 1470
            Fax: 519 457 3997
            Email: gijzenm@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Glycine max"
                     /cultivar="Harosoy 63"
                     /db_xref="taxon:3847"
                     /clone_lib="GmaxSC"
                     /tissue_type="Seed coats"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
                     cDNA library was constructed from polyA+ enriched mRNA
                     from  green seed coats in mid to late developemental stage
                     , average fresh  weight 250 mg per seed. Traces of pod and
                     embryo tissue also present.  Complementary DNA was
                     synthesized from mRNA using an XhoI-poly(dT)
                     linker-primer. EcoRI adopters were ligated to the
                     blunt-ended cDNA  fragments and the products were digested
                     with XhoI for directional  cloning into lambda ZAP Express
                     vector. This lambda library was  amplified once using E.
                     coli host strain XL1 Blue MRF'. Inserts were  then
                     subcloned by mass excision using ExAssist helper phage for
                     conversion into phagemid vector pBK-CMV in E. coli host
                     strain XLOLR."
BASE COUNT      199 a    147 c    148 g    203 t      6 others
ORIGIN

  Query Match             71.9%;  Score 53.2;  DB 10;  Length 703;
  Best Local Similarity   87.9%;  Pred. No. 3.4e-09;
  Matches   58;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTCA 61
           |||||||||||||||||| |||||||||||||||| |||||||||||||||| | |  |
Db     535 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCT 476

Qy      62 AACCGC 67
           | ||||
Db     475 AGCCGC 470


RESULT 3
B51582/c
LOCUS       B51582                   329 bp    DNA     linear   GSS 18-JUN-1998
DEFINITION  CIT978SK-A-481E4.TV CIT978SK Homo sapiens genomic clone A-481E4,
            DNA sequence.
ACCESSION   B51582
VERSION     B51582.1  GI:2603819
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 329)
  AUTHORS   Kim,U.-J., Adams,M.D. and Simon,M.I.
  TITLE     Determination of clone end sequences of human Bacterial Artificial
            Chromosomes
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT978SK-A-481E4.TP
            Contact: Ung-Jin Kim
            CalTech Genome Research Lab
            California Institute of Technology
            Division of Biology, MS 147-75, Pasadena, CA 91125, USA
            Tel: 626 796 7066
            Fax: 626 395 4901
            Email: ung@ash.tree.caltech.edu
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .329
                     /organism="Homo sapiens"
                     /db_xref="GDB:726891"
                     /db_xref="taxon:9606"
                     /clone="A-481E4"
                     /clone_lib="CIT978SK"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
                     CalTech Human BAC Library A"
BASE COUNT       72 a     89 c     90 g     78 t
ORIGIN

  Query Match             71.1%;  Score 52.6;  DB 17;  Length 329;
  Best Local Similarity   93.2%;  Pred. No. 4.2e-09;
  Matches   55;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           |||||||||||||||||| |||||||||||||||||||||||||||||||| | || |
Db     242 GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 184


RESULT 4
BQ248969/c
LOCUS       BQ248969                 388 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25001G10F TaE25 Triticum aestivum cDNA clone TaE25001G10F, mRNA
            sequence.
ACCESSION   BQ248969
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 558.039 Seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-3
Perfect score: 74
Sequence: 1 tggcgtgtacggtgggaggc.....accgtcaaaccgcggaagct 74

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	53.2	71.9	657	9	AI815377	AI815377 B0283860
c 2	53.2	71.9	703	10	BE661741	BE661741 138t7 Gma
c 3	52.6	71.1	329	17	B51582	B51582 CIT978SK-A-
c 4	52.6	71.1	388	14	BQ248969	BQ248969 TaE25001G
c 5	52.6	71.1	502	14	BQ248026	BQ248026 TaE25014A
c 6	52.6	71.1	535	14	BQ248278	BQ248278 TaE25011A
c 7	52.6	71.1	566	14	BQ248646	BQ248646 TaE25006C
c 8	52.6	71.1	575	14	BQ248555	BQ248555 TaE25007E
c 9	52.6	71.1	593	14	BQ248056	BQ248056 TaE25013F
c 10	52.6	71.1	610	17	B68191	B68191 CIT978SK-A-
c 11	52.6	71.1	625	14	BQ248255	BQ248255 TaE25011C
c 12	52.6	71.1	625	14	BQ248467	BQ248467 TaE25008F
c 13	52.6	71.1	630	14	BQ248453	BQ248453 TaE25008H
c 14	52.6	71.1	631	14	BQ248173	BQ248173 TaE25012C
c 15	52.6	71.1	633	14	BQ248939	BQ248939 TaE25002C
c 16	52.6	71.1	637	14	BQ248504	BQ248504 TaE25008B
c 17	52.6	71.1	640	14	BQ247936	BQ247936 TaE25015B
c 18	52.6	71.1	641	14	BQ248159	BQ248159 TaE25012E
c 19	52.6	71.1	642	14	BQ247761	BQ247761 TaE25042H
c 20	52.6	71.1	642	14	BQ247949	BQ247949 TaE25015A
c 21	52.6	71.1	644	14	BQ247963	BQ247963 TaE25014G
c 22	52.6	71.1	645	14	BQ247960	BQ247960 TaE25014H
c 23	52.6	71.1	646	14	BQ247762	BQ247762 TaE25042H
c 24	52.6	71.1	646	14	BQ248341	BQ248341 TaE25010C
c 25	52.6	71.1	647	14	BQ248311	BQ248311 TaE25010F
c 26	52.6	71.1	647	14	BQ248458	BQ248458 TaE25008G
c 27	52.6	71.1	648	14	BQ247997	BQ247997 TaE25014D
c 28	52.6	71.1	648	14	BQ248418	BQ248418 TaE25009C
c 29	52.6	71.1	648	14	BQ248913	BQ248913 TaE25002F
c 30	52.6	71.1	649	14	BQ247933	BQ247933 TaE25015B
c 31	52.6	71.1	649	14	BQ248366	BQ248366 TaE25009H
c 32	52.2	70.5	400	17	B63118	B63118 CIT978SK-A-
c 33	51	68.9	287	17	B63084	B63084 CIT978SK-A-
c 34	51	68.9	545	14	BQ248598	BQ248598 TaE25006H
c 35	51	68.9	550	14	BQ248595	BQ248595 TaE25006H
c 36	50.8	68.6	655	10	AW355220	AW355220 pnf-b.pk0
c 37	50.2	67.8	607	10	AV645401	AV645401 AV645401
c 38	50.2	67.8	617	10	AV645332	AV645332 AV645332
c 39	50.2	67.8	618	10	AV645383	AV645383 AV645383
c 40	50.2	67.8	619	10	AV645377	AV645377 AV645377
c 41	50.2	67.8	619	10	AV645379	AV645379 AV645379
c 42	50.2	67.8	619	10	AV645385	AV645385 AV645385
c 43	50.2	67.8	619	10	AV645389	AV645389 AV645389
c 44	50.2	67.8	620	10	AV645369	AV645369 AV645369
c 45	50.2	67.8	620	10	AV645380	AV645380 AV645380

ALIGNMENTS

RESULT 1
AI815377/c
LOCUS AI815377 657 bp mRNA linear EST 09-JUL-1999
DEFINITION B0283860 Zebrafish Kidney cDNA random primed, RZPD library no: 576
Danio rerio cDNA clone RZPD clone CHBOp576F21214Q3, mRNA sequence.
ACCESSION AI815377
VERSION AI815377.1 GI:5430923
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 657)
AUTHORS Look,A.T. and Holloway,M.
TITLE Zebrafish Kidney cDNA
JOURNAL Unpublished (1999)
COMMENT Contact: Thomas Look
Dana-Farber Cancer Institute
Pediatric Oncology Department, Mayer Building-630,; 44 Binney Street , Boston, MA 02115, USA
Tel: (617) 632-5826
Fax: (617) 632-6989
Email: thomas_look@dfci.harvard.edu
The clone was obtained from an RZPD array made from an adult kidney cDNA random primed library from Leonard Zon. RZPD library number 576. This clone is available from the Resource Centre/Primary Database of the German Human Genome Project: RZPD (Ressourcen Zentrum Primr Datenbank). http://www.rzpd.de

```
Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385
```

```
RESULT 15
AAX60047
ID   AAX60047 standard; DNA; 450 BP.
XX
AC   AAX60047;
XX
DT   04-AUG-1999  (first entry)
XX
DE   Minimal Human cytomegalovirus promoter.
XX
KW   Transgenic mice; transgene; tet operator-linked gene; tetracycline;
KW   mouse-active transcriptional regulatory element; mutant Tet repressor;
KW   gene therapy; genetic disease; acquired disease; cancer; viral disease;
KW   vaccination; rheumatoid arthritis; hypopituitarism; wound healing;
KW   tissue regeneration; cancer; benign prostatic hypertrophy; hemophilia;
KW   erythrocytopenia; artherosclerosis; liver disease; Alzheimer's disease;
KW   Parkinson's disease; human disease model; ds.

     Human cytomegalovirus.
XX
PN   US5912411-A.
XX
PD   15-JUN-1999.
XX
PF   07-JUN-1995;   93US-0487472.
XX
PR   07-JUN-1995;   95US-0487472.
PR   14-JUN-1993;   93US-0076327.
PR   14-JUN-1993;   93US-0076726.
PR   14-JUN-1994;   94US-0260452.
PR   01-JUL-1994;   94US-0270637.
PR   15-JUL-1994;   94US-0275876.
PR   03-FEB-1995;   95US-0383754.
XX
PA   (UYHE-) UNIV HEIDELBERG.
XX
PI   Bujard H,  Gossen M;
XX
DR   WPI; 1999-357232/30.
XX
PT   Transgenic mice with inducible transgene activity useful for in
PT   vitro and in vivo protein production
XX
PS   Disclosure; Column 71-72; 63pp; English.
XX
     The specification describes transgenic mice which have a transgene and a
     tet operator-linked gene integrated in the genome. The transgene
CC   comprises a mouse-active transcriptional regulatory element linked to a
CC   polynucleotide sequence that encodes a fusion protein which activates
CC   transcription of the tet operator-linked gene. The fusion protein
CC   comprises a mutated Tet repressor that binds a tet operator sequence
CC   in the presence of tetracycline (Tc) or a Tc analogue, linked to a
CC   polypeptide that activates transcription in eukaryotic cells. The
CC   transgenic system may be used for gene therapy to treat genes involved
CC   in genetic or acquired diseases. Gene therapy may be used to treat
CC   cancer, viral diseases, for vaccination, and to provide (Tc induced)
CC   regulated doses of a product (e.g. for the treatment or regulation  of
CC   rheumatoid arthritis, hypopituitarism, wound healing and tissue
CC   regeneration, cancer, benign prostatic hypertrophy, hemophilia,
CC   erythrocytopenia, artherosclerosis and liver disease, Alzheimer's
CC   disease, and Parkinson's disease). The system may also be used to
CC   produce proteins in vivo (e.g. using mammalian, yeast or fungal cells)
CC   or in vitro (e.g. transgenic farm animals), to produce animal models of
CC   human disease, or to produce a stable cell line for gene cloning. The
CC   present sequence is used to construct the transgenes of the invention.
XX
SQ   Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;
```

```
  Query Match           73.2%;  Score 54.2;  DB 20;  Length 450;
  Best Local Similarity 94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385
```

```
Search completed: March  5, 2003, 22:49:57
Job time : 73.2126 secs
```

```
CC   (tTA) integrated at a predetermined location within chromosomal
CC   DNA. The transgenic mice are useful for studying gene function and
CC   expression. The method allows the development of knockout animals for
CC   many genes of interest and allows for conditional inactivation or
CC   modulation of expression of a gene in a mouse.
XX
SQ   Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 20;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 13
AAX81722
ID   AAX81722 standard; DNA; 450 BP.
XX
AC   AAX81722;
XX
DT   27-AUG-1999  (first entry)
XX
DE   A human cytomegalovirus promoter sequence.
XX
KW   Tetracycline-controllable transactivator; tTA; transgenic mouse;
KW   transgene; gene function; gene expression; knockout animal;
KW   conditional inactivation; promoter; ds.
XX
OS   Human cytomegalovirus.
XX
PN   US5922927-A.
XX
PD   13-JUL-1999.
XX
PF   21-JUL-1997;   97US-0897719.
XX
PR   14-JUN-1994;   94US-0260452.
PR   14-JUN-1993;   93US-0076327.
PR   21-JUL-1997;   97US-0897719.
XX
PA   (BADI ) BASF AG.
XX
PI   Bujard H,  Gossen M,  Salfeld JG,  Voss JW;
XX
DR   WPI; 1999-404496/34.
XX
PT   Production of tetracycline-regulated transgenic mice, useful for
PT   studying gene expression
XX
PS   Disclosure; Fig 7; 66pp; English.
XX
CC   The present sequence represents a human cytomegalovirus promoter.
CC   It is used to create the transgenic animals of the invention. The
CC   specification describes a method for producing a transgenic mouse
CC   having a transgene encoding a tetracycline-controllable transactivator
CC   (tTA) integrated at a predetermined location within chromosomal
CC   DNA. The transgenic mice are useful for studying gene function and
CC   expression. The method allows the development of knockout animals for
CC   many genes of interest and allows for conditional inactivation or
CC   modulation of expression of a gene in a mouse.
XX
SQ   Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 20;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 14
AAX60046
ID   AAX60046 standard; DNA; 450 BP.
XX
AC   AAX60046;
XX
DT   04-AUG-1999  (first entry)
XX
DE   Minimal Human cytomegalovirus promoter.
XX
KW   Transgenic mice; transgene; tet operator-linked gene; tetracycline;
KW   mouse-active transcriptional regulatory element; mutant Tet repressor;
KW   gene therapy; genetic disease; acquired disease; cancer; viral disease;
KW   vaccination; rheumatoid arthritis; hypopituitarism; wound healing;
KW   tissue regeneration; cancer; benign prostatic hypertrophy; hemophilia;
KW   erythrocytopenia; artherosclerosis; liver disease; Alzheimer's disease;
KW   Parkinson's disease; human disease model; ds.
XX
OS   Human cytomegalovirus.
XX
PN   US5912411-A.
XX
PD   15-JUN-1999.
XX
PF   07-JUN-1995;   93US-0487472.
XX
PR   07-JUN-1995;   95US-0487472.
PR   14-JUN-1993;   93US-0076327.
PR   14-JUN-1993;   93US-0076726.
PR   14-JUN-1994;   94US-0260452.
PR   01-JUL-1994;   94US-0270637.
PR   15-JUL-1994;   94US-0275876.
PR   03-FEB-1995;   95US-0383754.
XX
PA   (UYHE-) UNIV HEIDELBERG.
XX
PI   Bujard H,  Gossen M;
XX
DR   WPI; 1999-357232/30.
XX
PT   Transgenic mice with inducible transgene activity useful for in
PT   vitro and in vivo protein production
XX
PS   Disclosure; Column 71-72; 63pp; English.
XX
CC   The specification describes transgenic mice which have a transgene and a
CC   tet operator-linked gene integrated in the genome. The transgene
CC   comprises a mouse-active transcriptional regulatory element linked to a
CC   polynucleotide sequence that encodes a fusion protein which activates
CC   transcription of the tet operator-linked gene. The fusion protein
CC   comprises a mutated Tet repressor that binds a tet operator sequence
CC   in the presence of tetracycline (Tc) or a Tc analogue, linked to a
CC   polypeptide that activates transcription in eukaryotic cells. The
CC   transgenic system may be used for gene therapy to treat genes involved
CC   in genetic or acquired diseases. Gene therapy may be used to treat
CC   cancer, viral diseases, for vaccination, and to provide (Tc induced)
CC   regulated doses of a product (e.g. for the treatment or regulation  of
CC   rheumatoid arthritis, hypopituitarism, wound healing and tissue
CC   regeneration, cancer, benign prostatic hypertrophy, hemophilia,
CC   erythrocytopenia, artherosclerosis and liver disease, Alzheimer's
CC   disease, and Parkinson's disease). The system may also be used to
CC   produce proteins in vivo (e.g. using mammalian, yeast or fungal cells)
CC   or in vitro (e.g. transgenic farm animals), to produce animal models of
CC   human disease, or to produce a stable cell line for gene cloning. The
CC   present sequence is used to construct the transgenes of the invention.
XX
SQ   Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 20;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

PI Bujard H, Gossen M;
XX
DR WPI; 1998-541795/46.
XX
PT Tetracycline based regulation of gene expression - uses a
PT tetracycline operator sequence joined to a gene of interest, the
PT gene of interest being induced in the presence, but not absence of
PT the antibiotic
XX
PS Example 6; Columns 71-72; 63pp; English.
XX
CC The present sequence represents a Cytomegalovirus (CMV) minimal promoter
CC linked to 10 tet operator sequences. The specification describes a
CC method for regulating expression of a Tet (tetracycline) operator-linked
CC gene in a cell of a subject. The method comprises introducing into the
CC cell a nucleic acid encoding a fusion protein which inhibits
CC transcription in eukaryotic cells, the fusion protein comprising a
CC polypeptide which binds to a Tet operator sequence, operatively linked
CC to heterologous second polypeptide which inhibits transcription in
CC eukaryotic cells and modulating the concentration of a tetracycline
CC (analogue) in the subject. The method is used for the regulation of
CC gene expression system, using tetracycline (analogues). The system
enables a gene coupled to the system to be induced in the presence of
Tet and then stopped when Tet is removed.
XX
SQ Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 19; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db 327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385

RESULT 11
AAV60081
ID AAV60081 standard; DNA; 450 BP.
XX
AC AAV60081;
XX
DT 04-DEC-1998 (first entry)
XX
DE Cytomegalovirus minimal promoter linked to 10 tet operator sequences.
XX
KW Tet repressor; tetracycline; regulation; expression;
KW Tet operator-linked gene; CMV minimal promoter; tet operator; ds.
XX
OS Synthetic.
OS Human cytomegalovirus.

US5814618-A.
XX
PD 29-SEP-1998.
XX
PF 07-JUN-1995; 95US-0485978.
XX
PR 07-JUN-1995; 95US-0485978.
PR 14-JUN-1993; 93US-0076327.
PR 14-JUN-1993; 93US-0076726.
PR 14-JUN-1994; 94US-0260452.
PR 01-JUL-1994; 94US-0270637.
PR 15-JUL-1994; 94US-0275876.
PR 06-FEB-1995; 95US-0383754.
XX
PA (BADI) BASF AG.
PA (KNOL) KNOLL AG.
XX
PI Bujard H, Gossen M;
XX
DR WPI; 1998-541795/46.
XX
PT Tetracycline based regulation of gene expression - uses a
PT tetracycline operator sequence joined to a gene of interest, the
PT gene of interest being induced in the presence, but not absence of
PT the antibiotic
XX
PS Disclosure; Columns 73-74; 63pp; English.
XX
CC The present sequence represents a Cytomegalovirus (CMV) minimal promoter
CC linked to 10 tet operator sequences. The specification describes a
CC method for regulating expression of a Tet (tetracycline) operator-linked
CC gene in a cell of a subject. The method comprises introducing into the
CC cell a nucleic acid encoding a fusion protein which inhibits
CC transcription in eukaryotic cells, the fusion protein comprising a
CC polypeptide which binds to a Tet operator sequence, operatively linked
CC to heterologous second polypeptide which inhibits transcription in
CC eukaryotic cells and modulating the concentration of a tetracycline
CC (analogue) in the subject. The method is used for the regulation of
CC gene expression system, using tetracycline (analogues). The system
CC enables a gene coupled to the system to be induced in the presence of
CC Tet and then stopped when Tet is removed.
XX
SQ Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 19; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db 327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385

RESULT 12
AAX81721
ID AAX81721 standard; DNA; 450 BP.
XX
AC AAX81721;
XX
DT 27-AUG-1999 (first entry)
XX
DE A human cytomegalovirus promoter sequence.
XX
KW Tetracycline-controllable transactivator; tTA; transgenic mouse;
KW transgene; gene function; gene expression; knockout animal;
KW conditional inactivation; promoter; ds.
XX
OS Human cytomegalovirus.
XX
PN US5922927-A.
XX
PD 13-JUL-1999.
XX
PF 21-JUL-1997; 97US-0897719.
XX
PR 14-JUN-1994; 94US-0260452.
PR 14-JUN-1993; 93US-0076327.
PR 21-JUL-1997; 97US-0897719.
XX
PA (BADI) BASF AG.
XX
PI Bujard H, Gossen M, Salfeld JG, Voss JW;
XX
DR WPI; 1999-404496/34.
XX
PT Production of tetracycline-regulated transgenic mice, useful for
PT studying gene expression
XX
PS Example 2; Fig 6; 66pp; English.
XX
CC The present sequence represents a human cytomegalovirus promoter.
CC It is used to create the transgenic animals of the invention. The
CC specification describes a method for producing a transgenic mouse
CC having a transgene encoding a tetracycline-controllable transactivator

PR 01-JUL-1994; 94US-0270637.
PR 15-JUL-1994; 94US-0275876.
PR 03-FEB-1995; 95US-0383754.
XX
PA (BUJA/) BUJARD H.
PA (GOSS/) GOSSEN M.
XX
PI Bujard H, Gossen M;
XX
DR WPI; 1996-087666/09.
XX
PT New tetracycline-regulated transcription modulators - comprising
PT fusion proteins which bind to tet operator sequences to activate or
PT inhibit transcription
XX
PS Disclosure; Page 73-74; 112pp; English.
XX
CC Fusion proteins comprising a first polypeptide which binds to a tet
CC operator sequence in the presence of tetracycline or a tetracycline
CC analogue, operatively linked to a second polypeptide which either
CC activates or inhibits transcription in eukaryotic cells may be used
CC to activate or inhibit transcription. Such proteins may be used to
CC regulate gene expression in cells and may be particularly useful for
CC gene therapy and for expression of gene products in transgenic
CC organisms. Induction of gene expression is rapid, efficient and
CC strong, typically 1000-2000 fold. The inducing agent does not cause
CC pleitropic effects or cytotoxicity in eukaryotic cells.
CC Alternatively, a self regulating construct encoding a transactivator
CC fusion protein can be created. Here, the fusion protein is
CC operatively linked to a minimal promoter also comprising at least
CC one tet operator sequence. This sequence is a CMV minimal promoter
CC and ten tet operator sequences. Other promoters are described in
CC AAT11356 and AAT11357.
XX
SQ Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 17; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db 327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385

RESULT 9
AAT11356
ID AAT11356 standard; DNA; 450 BP.
XX
AC AAT11356;
XX
DT 07-JUL-1996 (first entry)
XX
DE Minimal CMV promoter and ten tet operators.
XX
KW Tet repressor; Herpes simplex virus; HSV; virion protein 16; VP16;
KW fusion protein; gene expression; regulation; inhibition; activation;
KW transcription; ds.
XX
OS Transposon Tn-10/Cauliflower mosiac virus.
XX
PN WO9601313-A1.
XX
PD 18-JAN-1996.
XX
PF 29-JUN-1995; 95WO-US08179.
XX
PR 07-JUN-1995; 95US-0486814.
PR 01-JUL-1994; 94US-0270637.
PR 15-JUL-1994; 94US-0275876.
PR 03-FEB-1995; 95US-0383754.
XX
PA (BUJA/) BUJARD H.
PA (GOSS/) GOSSEN M.
XX
PI Bujard H, Gossen M;
XX
DR WPI; 1996-087666/09.
XX
PT New tetracycline-regulated transcription modulators - comprising
PT fusion proteins which bind to tet operator sequences to activate or
PT inhibit transcription
XX
PS Disclosure; Page 74; 112pp; English.
XX
CC Fusion proteins comprising a first polypeptide which binds to a tet
CC operator sequence in the presence of tetracycline or a tetracycline
CC analogue, operatively linked to a second polypeptide which either
CC activates or inhibits transcription in eukaryotic cells may be used
CC to activate or inhibit transcription. Such proteins may be used to
CC regulate gene expression in cells and may be particularly useful for
CC gene therapy and for expression of gene products in transgenic
CC organisms. Induction of gene expression is rapid, efficient and
CC strong, typically 1000-2000 fold. The inducing agent does not cause
CC pleitropic effects or cytotoxicity in eukaryotic cells.
CC Alternatively, a self regulating construct encoding a transactivator
CC fusion protein can be created. Here, the fusion protein is
CC operatively linked to a minimal promoter also comprising at least
CC one tet operator sequence. This sequence is a CMV minimal promoter
CC and ten tet operator sequences. Other promoters are described in
CC AAT11355 and AAT11357.
XX
SQ Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 17; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db 327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385

RESULT 10
AAV60080
ID AAV60080 standard; DNA; 450 BP.
XX
AC AAV60080;
XX
DT 04-DEC-1998 (first entry)
XX
DE Cytomegalovirus minimal promoter linked to 10 tet operator sequences.
XX
KW Tet repressor; tetracycline; regulation; expression;
KW Tet operator-linked gene; CMV minimal promoter; tet operator; ds.
XX
OS Synthetic.
OS Human cytomegalovirus.
XX
PN US5814618-A.
XX
PD 29-SEP-1998.
XX
PF 07-JUN-1995; 95US-0485978.
XX
PR 07-JUN-1995; 95US-0485978.
PR 14-JUN-1993; 93US-0076327.
PR 14-JUN-1993; 93US-0076726.
PR 14-JUN-1994; 94US-0260452.
PR 01-JUL-1994; 94US-0270637.
PR 15-JUL-1994; 94US-0275876.
PR 06-FEB-1995; 95US-0383754.
XX
PA (BADI) BASF AG.
PA (KNOL) KNOLL AG.
XX

```
Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 6
AAQ76266
ID   AAQ76266 standard; DNA; 450 BP.
XX
AC   AAQ76266;
XX
DT   17-JUL-1995  (first entry)
XX
DE   PhCMV*-1 promoter.
XX
KW   tTA; transactivator; tetracycline-controllable transactivator;
KW   conditional inactivation; homologous recombination; gene expression;
KW   gene regulation; gene therapy; tetracycline-resistance; tetR;
KW   transgenic animal; PhCMV*-1; promoter; tetO; CMV; ds.
XX
OS   Human cytomegalovirus K12, Towne.
XX
     Key             Location/Qualifiers
     mRNA            382..450
                     /*tag= a
XX
PN   WO9429442-A.
XX
PD   22-DEC-1994.
XX
PF   14-JUN-1994;   94WO-US06734.
XX
PR   14-JUN-1993;   93US-0076327.
XX
PA   (BADI ) BASF AG.
XX
PI   Bujard H,  Gossen M,  Salfeld JG,  Voss JW;
XX
DR   WPI; 1995-036472/05.
XX
PT   Regulatory systems using tetracycline-controllable transactivator
PT   (tTA)  - useful for conditional inactivation or modulation of
PT   gene expression in a host cell or animal
XX
PS   Disclosure; Page 52-53; 103pp; English.
XX
CC   The minimal promoter PhCMV* spans the human CMV intermediate-
CC   early promoter sequence from +75 to -53 (+1 being the first
CC   nucleotide transcribed).  TetO sequences were fused to
CC   this core promoter to give the new promoters PhCMV*-1 (given
CC   in AAQ76266) and PhCMV*-2 (AAQ76267).  These promoters are used
     to express tTA transactivator in host cells.

SQ   Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

  Query Match            73.2%;  Score 54.2;  DB 16;  Length 450;
  Best Local Similarity  94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 7
AAQ76267
ID   AAQ76267 standard; DNA; 450 BP.
XX
AC   AAQ76267;
XX
DT   17-JUL-1995  (first entry)
XX
DE   PhCMV*-2 promoter.
XX
KW   tTA; transactivator; tetracycline-controllable transactivator;
KW   conditional inactivation; homologous recombination; gene expression;
KW   gene regulation; gene therapy; tetracycline-resistance; tetR;
KW   transgenic animal; PhCMV*-2; promoter; tetO; CMV; ds.
XX
OS   Human cytomegalovirus Towne.
XX
FH   Key             Location/Qualifiers
FT   mRNA            382..450
FT                   /*tag= a
XX
PN   WO9429442-A.
XX
PD   22-DEC-1994.
XX
PF   14-JUN-1994;   94WO-US06734.
XX
PR   14-JUN-1993;   93US-0076327.
XX
PA   (BADI ) BASF AG.
XX
PI   Bujard H,  Gossen M,  Salfeld JG,  Voss JW;
XX
DR   WPI; 1995-036472/05.
XX
PT   Regulatory systems using tetracycline-controllable transactivator
PT   (tTA)  - useful for conditional inactivation or modulation of
PT   gene expression in a host cell or animal
XX
PS   Disclosure; Page 53; 103pp; English.
XX
CC   The minimal promoter PhCMV* spans the human CMV intermediate-
CC   early promoter sequence from +75 to -53 (+1 being the first
CC   nucleotide transcribed).  TetO sequences were fused to
CC   this core promoter to give the new promoters PhCMV*-1 (given
CC   in AAQ76266) and PhCMV*-2 (AAQ76267).  These promoters are used
CC   to express tTA transactivator in host cells.
XX
SQ   Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

  Query Match            73.2%;  Score 54.2;  DB 16;  Length 450;
  Best Local Similarity  94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 8
AAT11355
ID   AAT11355 standard; DNA; 450 BP.
XX
AC   AAT11355;
XX
DT   07-JUL-1996  (first entry)
XX
DE   Minimal CMV promoter and ten tet operators.
XX
KW   Tet repressor; Herpes simplex virus; HSV; virion protein 16; VP16;
KW   fusion protein; gene expression; regulation; inhibition; activation;
KW   transcription; ds.
XX
OS   Transposon Tn-10/Cauliflower mosiac virus.
XX
PN   WO9601313-A1.
XX
PD   18-JAN-1996.
XX
PF   29-JUN-1995;   95WO-US08179.
XX
PR   07-JUN-1995;   95US-0486814.
```

```
XX
PS   Claim 12; Page 172; 241pp; English.
XX
CC   The invention relates to a construct which allows animals to be bred in
CC   captivity but renders them infertile in the wild by allowing reversible
CC   control over fertility and reproduction. The construct comprises a native
CC   promoter, a blocking DNA sequence contoured for and designed to abrogate
CC   a crucial gene's function or to cause its mis-expression, and a genetic
CC   switch to regulate controlled expression/repression of the blocker/gene
CC   knockout. The construct is useful for preventing embryogenesis or
CC   gametogenesis in animals by stably transforming an animal cell with
CC   the construct by microinjection, transfection or infection, where the
CC   construct stably integrates into the genome by homologous recombination,
CC   and implanting the cell into a host organism, where a whole animal
CC   develops from the implanted cell. The present sequence is a repressible
CC   promoter comprising tet responsive element (TRE) which is coupled to and
CC   tightly regulates a minimal promoter region PminCMV from cytomegalovirus.
XX
SQ   Sequence 447 BP; 136 A; 104 C; 108 G; 99 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 22;  Length 447;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     321 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 379


RESULT 4
AAT06869
ID   AAT06869 standard; DNA; 450 BP.
XX
AC   AAT06869;
XX
DT   13-MAR-1996  (first entry)
XX
DE   PhCMV*-1 tetO construct.
XX
KW   Transactivator; tTA; tet operator; tetO; gene expression;
KW   tetracycline-responsive promoter; PhCMV*-1; human cytomegalovirus;
KW   transcription activation; eukaryotic cell; ds; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   TATA_signal     347..353
FT                   /*tag=  a
XX
PN   US5464758-A.
XX
PD   07-NOV-1995.
XX
PF   14-JUN-1993;   93US-0076726.
XX
PR   14-JUN-1993;   93US-0076726.
XX
PA   (BUJA/) BUJARD H.
PA   (GOSS/) GOSSEN M.
XX
PI   Bujard H,  Gossen M;
XX
DR   WPI; 1995-392612/50.
XX
PT   Polynucleotide encoding trans:activator fusion protein contg. tet
PT   repressor  - used to control expression of gene regulated by minimal
PT   promoter linked to tet operon, and vectors and cells where gene
PT   expression is regulated by tetracycline
XX
PS   Disclosure; Fig 6; 37pp; English.
XX
CC   The PhCMV*-1-tetO construct (AAT06869) is composed of tet operator
CC   sequences separated by 95 bases from a human cytomegalovirus minimal
CC   IE promoter.  A transactivator, tTA or tTAs (AAT06867-68), is used to
CC   regulate expression of a heterologous gene operably linked to
CC   PhCMV*-1-tetO.  On/off regulation of expression of the encoded
CC   heterologous protein by host eukaryotic cells is provided by varying
CC   the medium tetracycline conc.
XX
SQ   Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 16;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 5
AAT06870
ID   AAT06870 standard; DNA; 450 BP.
XX
AC   AAT06870;
XX
DT   13-MAR-1996  (first entry)
XX
DE   PhCMV*-2 tetO construct.
XX
KW   Transactivator; tTA; tet operator; tetO; gene expression;
KW   tetracycline-responsive promoter; PhCMV*-2; human cytomegalovirus;
KW   transcription activation; eukaryotic cell; ds; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   TATA_signal     347..353
FT                   /*tag=  a
XX
PN   US5464758-A.
XX
PD   07-NOV-1995.
XX
PF   14-JUN-1993;   93US-0076726.
XX
PR   14-JUN-1993;   93US-0076726.
XX
PA   (BUJA/) BUJARD H.
PA   (GOSS/) GOSSEN M.
XX
PI   Bujard H,  Gossen M;
XX
DR   WPI; 1995-392612/50.
XX
PT   Polynucleotide encoding trans:activator fusion protein contg. tet
PT   repressor  - used to control expression of gene regulated by minimal
PT   promoter linked to tet operon, and vectors and cells where gene
PT   expression is regulated by tetracycline
XX
PS   Disclosure; Fig 7; 37pp; English.
XX
CC   The PhCMV*-2-tetO construct (AAT06870) is composed of tet operator
CC   sequences separated by 76 bases from a human cytomegalovirus minimal
CC   IE promoter.  A transactivator, tTA or tTAs (AAT06867-68), is used to
CC   regulate expression of a heterologous gene operably linked to
CC   PhCMV*-2-tetO.  On/off regulation of expression of the encoded
CC   heterologous protein by host eukaryotic cells is provided by varying
CC   the medium tetracycline conc.
XX
SQ   Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

  Query Match             73.2%;  Score 54.2;  DB 16;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 4.3e-12;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

PT mammalian gene expression, is based on the interferon regulating
PT factor-1 and its binding site -
XX
PS Claim 1; Page 9; 19pp; English.
XX
CC The present invention describes a promoter-transactivator system for
CC inducible high-level expression of mammalian genes, and optionally
CC for control of cell growth. The promoter-transactivator system
CC comprises: (i) a promoter construct (IRFE promoter); and (ii)
CC transactivator construct encoding a fusion protein of IRF-1 (interferon
CC regulating factor-1) and the oestrogen receptor (ER). The IRFE promoter
CC construct has the structure (MPSV-E)-(IRF-1 binding site)-(CMV)-DNA
CC where MPSV-E indicates the myeloproliferative sarcoma virus enhancer
CC repeat given in AAC64591; IRF-1 binding site given in AAC64592; and CMV
CC is the cytomegalovirus minimal promoter given in AAC64593 or their
CC functionally equivalent variants with one or more nucleotides
CC substituted, inserted or deleted. The promoter transactivator system is
CC a transcription regulator. Increased transcription results from binding
CC of the IRFE promoter to the transactivator, which is activated by
CC oestradiol or other ER ligands (these compounds displace the heat-shock
CC protein 90 which normally binds to the IRF-1/ER fusion, preventing its
CC activation). The system is used for increasing production of
CC therapeutically active proteins and where a IRF-1-green fluorescent
CC protein-human ER construct is used, for fluorescence-activated cell
CC sorting/analysis of transformed cells.
XX
SQ Sequence 74 BP; 19 A; 17 C; 23 G; 15 T; 0 other;

Query Match 100.0%; Score 74; DB 21; Length 74;
Best Local Similarity 100.0%; Pred. No. 2.5e-20;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60

Qy     61 AAACCGCGGAAGCT 74
          ||||||||||||||
Db     61 AAACCGCGGAAGCT 74
```

RESULT 2
AAT47525
ID AAT47525 standard; DNA; 432 BP.
XX
AC AAT47525;
XX
DT 08-AUG-1997 (first entry)
XX
DE tetO/eCMV chimeric target sequence encoding DNA.
XX
KW Tetracycline; transactivator; anti-hepadnaviral agent; ss.
XX
OS Chimeric - Human cytomegalovirus.
OS Chimeric - Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_signal 1..312
FT /*tag= a
FT /function= operator
FT /organism= Escherichia coli
FT promoter 313..432
FT /*tag= b
FT /organism= Human cytomegalovirus
XX
PN WO9639542-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-US08772.
XX
PR 05-JUN-1995; 95US-0462216.
XX
PA (AVID-) AVID THERAPEUTICS.
XX
PI Barker CS; King RW, Seeger C;
XX
DR WPI; 1997-043163/04.
XX
PT Novel cell lines for inducing expression of hepatitis B virus genome
PT - used for cell-based assay to screen potential anti-hepadnaviral
PT agents
XX
PS Claim 8; Page 36; 49pp; English.
XX
CC A novel cell line capable of inducibly expressing a genome of
CC hepatitis B virus (HBV), is transformed with a HBV genome operably
CC linked to a target nucleotide sequence for activating expression of the
CC HBV genome, the activation of expression being caused by interaction of
CC the target sequence with a transactivator protein that specifically
CC interacts with the target sequence. The present sequence encodes a
CC target sequence which contains a prokaryotic tetracycline operator
CC (tetO) sequence fused to human cytomegalovirus promoter IE (hCMV).
CC The promoter is activated by the tetR/VP16 chimeric transactivator
CC protein (E.coli tet repressor fused to the activating domain of virion
CC protein 16 of herpes simplex virus). The method and cell lines can be
CC used for a cell-based assay to screen potential anti-hepadnaviral
CC agents.
XX
SQ Sequence 432 BP; 135 A; 94 C; 105 G; 98 T; 0 other;

Query Match 73.2%; Score 54.2; DB 18; Length 432;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    315 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 373
```

RESULT 3
AAD09962
ID AAD09962 standard; DNA; 447 BP.
XX
AC AAD09962;
XX
DT 12-SEP-2001 (first entry)
XX
DE Tet responsive element (TRE)-PminCMV repressible promoter.
XX
KW Tet responsive element; TRE; promoter; reproduction; embryogenesis; CMV;
KW cytomegalovirus; gametogenesis; microinjection; fertility; infection;
KW chimeric; ds.
XX
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200148224-A1.
XX
PD 05-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-AU01596.
XX
PR 24-DEC-1999; 99AU-0004884.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;
PI Patil J;
XX
DR WPI; 2001-425672/45.
XX
PT Novel construct for preventing embryogenesis in animals comprises
PT native promoter, blocking DNA which abrogates function of crucial gene
PT and genetic switch to regulate expression/repression of blocker/gene
PT knockout -

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:12 ; Search time 72.2126 Seconds
(without alignments)
2307.741 Million cell updates/sec

Title: US-09-980-548-3
Perfect score: 74
Sequence: 1 tggcgtgtacggtgggaggc.....accgtcaaaccgcggaagct 74

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	74	21	AAC64593	Cytomegalovirus mi
2	54.2	73.2	432	18	AAT47525	tetO/eCMV chimeric
3	54.2	73.2	447	22	AAD09962	Tet responsive ele
4	54.2	73.2	450	16	AAT06869	PhCMV*-1 tetO cons
5	54.2	73.2	450	16	AAT06870	PhCMV*-2 tetO cons
6	54.2	73.2	450	16	AAQ76266	PhCMV*-1 promoter.
7	54.2	73.2	450	16	AAQ76267	PhCMV*-2 promoter.
8	54.2	73.2	450	17	AAT11355	Minimal CMV promot
9	54.2	73.2	450	17	AAT11356	Minimal CMV promot
10	54.2	73.2	450	19	AAV60080	Cytomegalovirus mi
11	54.2	73.2	450	19	AAV60081	Cytomegalovirus mi
12	54.2	73.2	450	20	AAX81721	A human cytomegalo
13	54.2	73.2	450	20	AAX81722	A human cytomegalo
14	54.2	73.2	450	20	AAX60046	Minimal Human cyto
15	54.2	73.2	450	20	AAX60047	Minimal Human cyto
16	54.2	73.2	450	20	AAX27901	PhCMV*-1 promoter
17	54.2	73.2	450	20	AAX27903	PhCMV*-2 promoter
18	54.2	73.2	450	20	AAX01365	PhCMV*-1 sequence.
19	54.2	73.2	450	20	AAX01367	PhCMV*-2 sequence.
20	54.2	73.2	450	21	AAZ56127	Tetracycline respo
21	54.2	73.2	450	21	AAZ56128	Tetracycline respo
22	54.2	73.2	450	22	AAD09834	Human cytomegalovi
23	54.2	73.2	450	22	AAD09835	Human cytomegalovi
24	54.2	73.2	450	22	AAH25573	Nucleotide sequenc
25	54.2	73.2	450	22	AAH25574	Nucleotide sequenc
26	54.2	73.2	450	22	AAH47632	Tetracycline-respo
27	54.2	73.2	450	22	AAH47633	Nucleotide sequenc
28	54.2	73.2	470	20	AAX90500	Tet promoter clone
29	54.2	73.2	481	19	AAV17755	Inducible promoter
30	54.2	73.2	481	19	AAV18691	Cytomegalovirus im
31	54.2	73.2	500	24	ABL41162	TetO/minimal CMV t
32	54.2	73.2	520	17	AAT11359	Bidirectional prom
33	54.2	73.2	520	18	AAT45722	Bi-directional tet
34	54.2	73.2	520	19	AAV60079	Bidirectional prom
35	54.2	73.2	520	20	AAX60045	Bidirectional prom
36	54.2	73.2	520	21	AAZ56126	Bidirectional prom
37	54.2	73.2	520	22	AAH25572	Nucleotide sequenc
38	54.2	73.2	520	22	AAH47631	Nucleotide sequenc
39	54.2	73.2	569	17	AAT11358	Bidirectional prom
40	54.2	73.2	569	18	AAT45721	Bi-directional tet
41	54.2	73.2	569	19	AAV60078	Bidirectional prom
42	54.2	73.2	569	20	AAX60044	Bidirectional prom
43	54.2	73.2	569	21	AAZ56125	Bidirectional prom
44	54.2	73.2	569	22	AAH25571	Nucleotide sequenc
45	54.2	73.2	569	22	AAH47630	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAC64593
ID AAC64593 standard; DNA; 74 BP.
XX
AC AAC64593;
XX
DT 15-FEB-2001 (first entry)
XX
DE Cytomegalovirus minimal promoter sequence SEQ ID NO:1.
XX
KW Myeloproliferative sarcoma virus; MPSV; MPSV-E; enhancer; CMV;
KW cytomegalovirus; firefly; IRF-1 binding site; minimal promoter;
KW interferon regulatory factor 1 binding site; transcription regulator;
KW promoter transactivator system; ds.
XX
OS Cytomegalovirus.
XX
PN EP1046710-A1.
XX
PD 25-OCT-2000.
XX
PF 23-APR-1999; 99EP-0108068.
XX
PR 23-APR-1999; 99EP-0108068.
XX
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI Mueller PP, Geserick C, Schroeder K, Hauser H;
XX
DR WPI; 2000-648930/63.
XX
PT Promoter-transactivator system, useful for inducing high level

```
VERSION     AR072128.1  GI:7223016
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Mice transgenic for a tetracycline-inducible transcriptional
            activator
  JOURNAL   Patent: US 5912411-A 8 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 12
AR072129
LOCUS       AR072129                450 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 9 from patent US 5912411.
ACCESSION   AR072129
VERSION     AR072129.1  GI:7223017
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Mice transgenic for a tetracycline-inducible transcriptional
            activator
  JOURNAL   Patent: US 5912411-A 9 15-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 13
AR095982
LOCUS       AR095982                450 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 8 from patent US 6004941.
ACCESSION   AR095982
VERSION     AR095982.1  GI:10024370
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Methods for regulating gene expression
  JOURNAL   Patent: US 6004941-A 8 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 14
AR095983
LOCUS       AR095983                450 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 9 from patent US 6004941.
ACCESSION   AR095983
VERSION     AR095983.1  GI:10024372
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Methods for regulating gene expression
  JOURNAL   Patent: US 6004941-A 9 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 15
AR136642
LOCUS       AR136642                450 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 8 from patent US 6136954.
ACCESSION   AR136642
VERSION     AR136642.1  GI:14477314
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Tetracycline-inducible transcriptional activator fusion proteins
  JOURNAL   Patent: US 6136954-A 8 24-OCT-2000;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db     327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


Search completed: March  5, 2003, 23:34:39
Job time : 408.715 secs
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H., Gossen,M., Salfeld,J.G. and Voss,J.W.
  TITLE     Mice transgenic for a tetracycline-controlled transcriptional
            activator
  JOURNAL   Patent: US 5859310-A 6 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 7
AR032151
LOCUS       AR032151                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 8 from patent US 5866755.
ACCESSION   AR032151
VERSION     AR032151.1  GI:5946440
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Animals transgenic for a tetracycline-regulated transcriptional
            inhibitor
  JOURNAL   Patent: US 5866755-A 8 02-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 8
AR032152
LOCUS       AR032152                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 9 from patent US 5866755.
ACCESSION   AR032152
VERSION     AR032152.1  GI:5946441
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Animals transgenic for a tetracycline-regulated transcriptional
            inhibitor
  JOURNAL   Patent: US 5866755-A 9 02-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 9
AR043816
LOCUS       AR043816                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 8 from patent US 5814618.
ACCESSION   AR043816
VERSION     AR043816.1  GI:5964824
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Methods for regulating gene expression
  JOURNAL   Patent: US 5814618-A 8 29-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 10
AR043817
LOCUS       AR043817                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 9 from patent US 5814618.
ACCESSION   AR043817
VERSION     AR043817.1  GI:5964825
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Methods for regulating gene expression
  JOURNAL   Patent: US 5814618-A 9 29-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match             73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity   94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 11
AR072128
LOCUS       AR072128                450 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 8 from patent US 5912411.
ACCESSION   AR072128
```

```
          Gesellschaft f}r Biotechnologische Forschung mbH (GBF) (; DE)
FEATURES             Location/Qualifiers
     source          1. .74
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       19 a     17 c     23 g     15 t
ORIGIN

  Query Match            100.0%;  Score 74;  DB 6;  Length 74;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-19;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60

Qy     61 AAACCGCGGAAGCT 74
          ||||||||||||||
Db     61 AAACCGCGGAAGCT 74


RESULT 2
  ?986
   US       I89986                  432 bp    DNA     linear   PAT 10-AUG-1998
DEFINITION  Sequence 1 from patent US 5723319.
ACCESSION   I89986
VERSION     I89986.1  GI:3409926
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   King,R.W., Barker,C.S. and Seeger,C.
  TITLE     Cultured cell line that inducibly expresses the hepatitis B virus
            genome, and uses thereof for screening antiviral substances
  JOURNAL   Patent: US 5723319-A 1 03-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="unknown"
BASE COUNT      135 a     94 c    105 g     98 t
ORIGIN

  Query Match            73.2%;  Score 54.2;  DB 6;  Length 432;
  Best Local Similarity  94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    315 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 373


   ULT 3
  ,20199
LOCUS       AR020199                450 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 8 from patent US 5789156.
ACCESSION   AR020199
VERSION     AR020199.1  GI:3974814
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Tetracycline-regulated transcriptional inhibitors
  JOURNAL   Patent: US 5789156-A 8 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match            73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity  94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 4
AR020200
LOCUS       AR020200                450 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 9 from patent US 5789156.
ACCESSION   AR020200
VERSION     AR020200.1  GI:3974815
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H. and Gossen,M.
  TITLE     Tetracycline-regulated transcriptional inhibitors
  JOURNAL   Patent: US 5789156-A 9 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      104 a    109 c    102 g    135 t
ORIGIN

  Query Match            73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity  94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 5
AR029414
LOCUS       AR029414                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 5 from patent US 5859310.
ACCESSION   AR029414
VERSION     AR029414.1  GI:5941387
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Bujard,H., Gossen,M., Salfeld,J.G. and Voss,J.W.
  TITLE     Mice transgenic for a tetracycline-controlled transcriptional
            activator
  JOURNAL   Patent: US 5859310-A 5 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="unknown"
BASE COUNT      138 a    101 c    110 g    101 t
ORIGIN

  Query Match            73.2%;  Score 54.2;  DB 6;  Length 450;
  Best Local Similarity  94.9%;  Pred. No. 8.5e-11;
  Matches   56;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAACCGTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||| | || |
Db    327 GGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC 385


RESULT 6
AR029415
LOCUS       AR029415                450 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 6 from patent US 5859310.
ACCESSION   AR029415
VERSION     AR029415.1  GI:5941388
KEYWORDS    .
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:16 ; Search time 407.715 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: US-09-980-548-3
Perfect score: 74
Sequence: 1 tggcgtgtacggtgggaggc.....accgtcaaaccgcggaagct 74

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	74	6	AX040911	AX040911 Sequence
2	54.2	73.2	432	6	I89986	I89986 Sequence 1
3	54.2	73.2	450	6	AR020199	AR020199 Sequence
4	54.2	73.2	450	6	AR020200	AR020200 Sequence
5	54.2	73.2	450	6	AR029414	AR029414 Sequence
6	54.2	73.2	450	6	AR029415	AR029415 Sequence
7	54.2	73.2	450	6	AR032151	AR032151 Sequence
8	54.2	73.2	450	6	AR032152	AR032152 Sequence
9	54.2	73.2	450	6	AR043816	AR043816 Sequence
10	54.2	73.2	450	6	AR043817	AR043817 Sequence
11	54.2	73.2	450	6	AR072128	AR072128 Sequence
12	54.2	73.2	450	6	AR072129	AR072129 Sequence
13	54.2	73.2	450	6	AR095982	AR095982 Sequence
14	54.2	73.2	450	6	AR095983	AR095983 Sequence
15	54.2	73.2	450	6	AR136642	AR136642 Sequence
16	54.2	73.2	450	6	AR136643	AR136643 Sequence
17	54.2	73.2	450	6	AR157177	AR157177 Sequence
18	54.2	73.2	450	6	AR157178	AR157178 Sequence
19	54.2	73.2	450	6	AR164126	AR164126 Sequence
20	54.2	73.2	450	6	AR164127	AR164127 Sequence
21	54.2	73.2	450	6	AX127249	AX127249 Sequence
22	54.2	73.2	450	6	AX127250	AX127250 Sequence
23	54.2	73.2	450	6	I15364	I15364 Sequence 12
24	54.2	73.2	450	6	I15365	I15365 Sequence 13
25	54.2	73.2	450	6	I32812	I32812 Sequence 8
26	54.2	73.2	450	6	I32813	I32813 Sequence 9
27	54.2	73.2	450	6	I56756	I56756 Sequence 5
28	54.2	73.2	450	6	I56757	I56757 Sequence 6
29	54.2	73.2	450	6	I59626	I59626 Sequence 8
30	54.2	73.2	450	6	I59627	I59627 Sequence 9
31	54.2	73.2	481	6	AR001287	AR001287 Sequence
32	54.2	73.2	481	6	AR116269	AR116269 Sequence
33	54.2	73.2	520	6	AR020198	AR020198 Sequence
34	54.2	73.2	520	6	AR032150	AR032150 Sequence
35	54.2	73.2	520	6	AR043815	AR043815 Sequence
36	54.2	73.2	520	6	AR072127	AR072127 Sequence
37	54.2	73.2	520	6	AR095981	AR095981 Sequence
38	54.2	73.2	520	6	AR136641	AR136641 Sequence
39	54.2	73.2	520	6	AR157176	AR157176 Sequence
40	54.2	73.2	520	6	AR164125	AR164125 Sequence
41	54.2	73.2	520	6	AX127248	AX127248 Sequence
42	54.2	73.2	520	6	I32811	I32811 Sequence 7
43	54.2	73.2	520	6	I59625	I59625 Sequence 7
44	54.2	73.2	569	6	AR020197	AR020197 Sequence
45	54.2	73.2	569	6	AR032149	AR032149 Sequence

ALIGNMENTS

RESULT 1
AX040911
LOCUS AX040911 74 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 3 from Patent WO0065074.
ACCESSION AX040911
VERSION AX040911.1 GI:11340533
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 74)
AUTHORS Mueller,P., Geserick,C., Schroeder,K. and Hauser,H.
TITLE Promoter-transactivator system for inducible high-level mammalian gene expression with the option of cell growth control
JOURNAL Patent: WO 0065074-A 3 02-NOV-2000;

KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 747)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: de36c01.x1
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by A.M. Zorn (Wellcome/CRC Institute). DNA
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Putative full length read
vector to vector length is
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 513.
FEATURES Location/Qualifiers
source 1. .747
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3473665"
/clone_lib="Wellcome CRC pRN3 dorsal lip"
/tissue_type="dorsal lip"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library was constructed
by A.M. Zorn (Wellcome/CRC Institute). "
BASE COUNT 177 a 201 c 195 g 172 t 2 others
ORIGIN

Query Match 39.1%; Score 27.4; DB 13; Length 747;
Best Local Similarity 65.6%; Pred. No. 1.5e+02;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
           | | |||||   |||  ||| | | |    |||| | || |||| |||| ||||| || |
         418 GCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGTTCCCTCTCCGGAATTGGAACCTTA 359

Qy      61 A 61
           |
Db     358 A 358
```

Search completed: March 6, 2003, 00:22:21
Job time : 533.208 secs

```
  Query Match             39.4%;  Score 27.6;  DB 14;  Length 753;
  Best Local Similarity   67.2%;  Pred. No. 1.3e+02;
  Matches   39;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      12 GGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           | ||   ||||| |||   ||||| | ||    |  ||| ||| ||||||| ||||||
Db     452 GCGACGAGGAAAGTGATGCTCAGAGCTCTGGAAGCAAAAGGGAGACTGAAATTCAGAT 509


RESULT 13
AV096262
LOCUS       AV096262                279 bp    mRNA    linear   EST 22-NOV-1999
DEFINITION  AV096262 Mus musculus C57BL/6J ES cell Mus musculus cDNA clone
            2410017M04, mRNA sequence.
ACCESSION   AV096262
VERSION     AV096262.1  GI:5243810
 ¬YWORDS    EST.
  URCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
            ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
            Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
            Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
            ,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
            Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
            Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
 ATURES             Location/Qualifiers
     source         1. .279
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="2410017M04"
                    /clone_lib="Mus musculus C57BL/6J ES cell"
                    /cell_type="ES cell"
                    /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                    ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                    RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                    primer {5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                    T 3'}; double-stranded cDNA was ligated to Eco RI
                    adaptors (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of the modified pT7T3 vector.
                    RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                    constructed and normalized by Bento Soares and M.Patima
                    Bonaldo."
BASE COUNT       90 a     54 c     61 g     74 t
ORIGIN

  Query Match             39.1%;  Score 27.4;  DB 9;  Length 279;
  Best Local Similarity   69.8%;  Pred. No. 1.3e+02;
  Matches   37;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      12 GGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAAT 64
           ||||||||||| ||  || |   |||| ||  | ||||||||  ||   ||||
Db       5 GGGAAATGGAACCTATAACTTCCATCCATTTACAGGAAATGGTGACCAGAAAT 57


RESULT 14
AQ664260
LOCUS       AQ664260                451 bp    DNA     linear   GSS 23-JUN-1999
DEFINITION  HS_5496_B1_F11_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1072 Col=21 Row=L, DNA sequence.
ACCESSION   AQ664260
VERSION     AQ664260.1  GI:5171951
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu                ;
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1072  row: L  column: 21
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 451.
FEATURES            Location/Qualifiers
     source         1. .451
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate=1072 Col=21 Row=L"
                    /clone_lib="RPCI-11 Human Male BAC Library"
                    /sex="male"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    Male blood DNA was isolated from one randomly chosen donor
                    and partially digested with a combination of EcoRI and
                    EcoRI Methylase. Size selected DNA was cloned into the
                    pBACe3.6 vector at EcoRI sites"
BASE COUNT      121 a    119 c    100 g    111 t
ORIGIN

  Query Match             39.1%;  Score 27.4;  DB 17;  Length 451;
  Best Local Similarity   69.8%;  Pred. No. 1.4e+02;
  Matches   37;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       2 ATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGA 54
           || |||||||   |||||| ||  |||| |  |||  ||||||  | | ||||
Db     123 ATGCCTTCTCACTAAATGGCAAAGGAAAGTAGGATTTCTTCTCTTGGATTGGA 175


RESULT 15
BI938198/c
LOCUS       BI938198                747 bp    mRNA    linear   EST 18-OCT-2001
DEFINITION  de36c01.y1 Wellcome CRC pRN3 dorsal lip Xenopus laevis cDNA clone
            IMAGE:3473665 5' similar to TR:O14597 O14597 NON-FUNCTIONAL FOLATE
            BINDING PROTEIN. ;, mRNA sequence.
ACCESSION   BI938198
VERSION     BI938198.1  GI:16252670
```

```
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .656
                      /organism="Triticum aestivum"
                      /cultivar="Chinese Spring"
                      /db_xref="taxon:4565"
                      /clone="whh2n20"
                      /clone_lib="Y. Ogihara unpublished cDNA library, Wh_h"
                      /tissue_type="spike at heading date"
                      /dev_stage="Feekes' scale 10.5"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                      Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                      hydroponic conditions at UC Davis, salt stressed for 12
                      hours, and for 7 days, then dissected and frozen (Akhunov
                      in J Dvorak Lab). Total RNA was prepared from sheath
                      tissue, equal quantities of RNA were pooled from the two
                      samples, polyA was purified from the pooled RNA, a cDNA
                      library was made, and the cDNA clones were in vivo
                      excised to give pBluescript phagemids in the TJ Close lab
                      at the University of California, Riverside (Akhunov, Chin
                      , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
                      Plasmid DNA preparations and DNA sequencing were
                      performed in the OD Anderson lab (all other authors)."
BASE COUNT      213 a    137 c    160 g    146 t
ORIGIN

  Query Match             39.4%;  Score 27.6;  DB 13;  Length 656;
  Best Local Similarity   67.2%;  Pred. No. 1.3e+02;
  Matches   39;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      12 GGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           | ||   ||||| |||   ||||| | ||    |  ||| ||| ||||||| ||||||
Db     131 GCGACGAGGAAAGTGATGCTCAGAGCTCTGGAAGCAAAAGGGAGACTGAAATTCAGAT 188


RESULT 11
BQ238956
LOCUS       BQ238956                658 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE05039B10R TaE05 Triticum aestivum cDNA clone TaE05039B10R, mRNA
            sequence.
ACCESSION   BQ238956
VERSION     BQ238956.1  GI:20434832
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
  :RENCE    1  (bases 1 to 658)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >2.0 kb
            Plate: 039  row: B  column: 10
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .658
                      /organism="Triticum aestivum"
                      /cultivar="Glenlea"
                      /db_xref="taxon:4565"
                      /clone="TaE05039B10R"
                      /clone_lib="TaE05"
                      /tissue_type="developing seeds"
                      /dev_stage="5 days after anthesis"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
                      NotI; Site_2: MluI; mRNA obtained from wheat seeds of
                      cultivar Glenlea 5 days post-anthesis"
BASE COUNT      204 a    111 c    150 g    193 t
ORIGIN

  Query Match             39.4%;  Score 27.6;  DB 14;  Length 658;
  Best Local Similarity   67.2%;  Pred. No. 1.3e+02;
  Matches   39;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      12 GGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           | ||   ||||| |||   ||||| | ||    |  ||| ||| ||||||| ||||||
Db     324 GCGACGAGGAAAGTGATGCTCAGAGCTCTGGAAGCAAAAGGGAGACTGAAATTCAGAT 381


RESULT 12
BQ842141
LOCUS       BQ842141                753 bp    mRNA    linear   EST 12-AUG-2002
DEFINITION  WHE2989_D07_H13ZS Wheat dormant embryo cDNA library Triticum
            aestivum cDNA clone WHE2989_D07_H13, mRNA sequence.
ACCESSION   BQ842141
VERSION     BQ842141.1  GI:22211550
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Anderson,O.D., Chao,S., Chin,A., Close,T.J., Doherty,L., Fenton
            ,R.D., Lazo,G.R., Rausch,C.J., Walker-Simmons,M.K. and Wilson,C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Dormant embryo cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES              Location/Qualifiers
     source           1. .753
                      /organism="Triticum aestivum"
                      /cultivar="Brevor"
                      /db_xref="taxon:4565"
                      /clone="WHE2989_D07_H13"
                      /clone_lib="Wheat dormant embryo cDNA library"
                      /tissue_type="Seed embryo"
                      /dev_stage="Mature seed"
                      /lab_host="E. coli SOLR"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                      Site_1: EcoRI; Site_2: XhoI; Plants were grown to seed
                      maturity under conditions favoring seed dormancy (L.
                      Dohery at K. Walker_Simmons lab, Washington State
                      University, Pullman, WA). Embryos were cut from mature
                      dormant seed (Doherty). Total RNA was prepared from these
                      embryos, polyA was purified, a cDNA library was made, and
                      the cDNA clones were in vivo excised to give pBluescript
                      phagemids in the TJ Close lab at the University of
                      California, Riverside (Chin, Fenton). Plasmid DNA
                      preparations and DNA sequencing were performed in the OD
                      Anderson lab (all other authors)."
BASE COUNT      216 a    150 c    202 g    185 t
ORIGIN
```

```
LOCUS       T97467                 434 bp    mRNA    linear   EST 29-MAR-1995
DEFINITION  ye53b12.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:121439 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   T97467
VERSION     T97467.1  GI:746812
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
  COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1020
            Source: IMAGE Consortium, LLNL This clone is available royalty-free
            through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov)
            for further information.
            Insert Length: 1020   Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 435.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /db_xref="GDB:473984"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:121439"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      103 a     98 c     89 g    143 t      1 others
ORIGIN

  Query Match             39.7%;  Score 27.8;  DB 14;  Length 434;
  Best Local Similarity   65.1%;  Pred. No. 1e+02;
  Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       7 TTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCA 66
           | | | |  | || ||    ||||| ||| ||||  | | |||| |||||    ||||||
Db     366 TACCCTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCA 307

Qy      67 GAT 69
           |||
Db     306 GAT 304


RESULT 9
AQ630605
LOCUS       AQ630605               502 bp    DNA     linear   GSS 17-JUN-1999
DEFINITION  RPCI-11-484I10.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-484I10
            , DNA sequence.
ACCESSION   AQ630605
VERSION     AQ630605.1  GI:5093240
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
  COMMENT   Other_GSSs: RPCI-11-484I10.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="GDB:7685673"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-484I10"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      187 a     75 c    103 g    137 t
ORIGIN

  Query Match             39.4%;  Score 27.6;  DB 17;  Length 502;
  Best Local Similarity   63.6%;  Pred. No. 1.2e+02;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       5 CCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAAT 64
           || || | | ||||||||||||||  ||  | | ||| ||   | ||||  |  ||||
Db     105 CCATCACTGAAAATGGAAACTGATCATTTGGTGCCTCCTGATGTAATGTGATATGAAGCA 164

Qy      65 CAGATC 70
           || | |
Db     165 CACAGC 170


RESULT 10
BJ255893
LOCUS       BJ255893               656 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BJ255893 Y. Ogihara unpublished cDNA library, Wh_h Triticum
            aestivum cDNA clone whh2n20 5', mRNA sequence.
ACCESSION   BJ255893
VERSION     BJ255893.1  GI:20079675
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
```

```
ORIGIN

  Query Match             40.3%;  Score 28.2;  DB 13;  Length 391;
  Best Local Similarity   68.4%;  Pred. No. 77;
  Matches   39;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      5 CCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAA 61
          |||||| | ||  ||| | | |   |||||| || |||| ||||| | ||  ||||
Db    339 CCTTCTTGAGATGTGGTAGCCGTTTATCAGACTCCCTCTCCGGAAACGAAACATGAA 283


RESULT 6
AZ913831/c
LOCUS       AZ913831                 540 bp    DNA     linear   GSS 05-MAR-2001
DEFINITION  RPCI-24-174P15.TV RPCI-24 Mus musculus genomic clone RPCI-24-174P15
            , DNA sequence.
ACCESSION   AZ913831
VERSION     AZ913831.1  GI:13232776
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE 1  (bases 1 to 540)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-174P15.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 174  row: P  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-174P15"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      179 a    106 c    112 g    143 t
ORIGIN

  Query Match             40.3%;  Score 28.2;  DB 17;  Length 540;
  Best Local Similarity   64.6%;  Pred. No. 82;
  Matches   42;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      4 CCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAA 63
          || ||||    |||||| ||||||   |  | | || |||| || | ||||||   | ||
Db    164 CCTTTCTTCCAAAATGGGAACTGTGTACGAAAGCCTTTCTTGGAACATGGAATGAGGAAT 105

Qy     64 TCAGA 68
          ||||
Db    104 GCAGA 100
```

```
RESULT 7
R94003/c
LOCUS       R94003                   391 bp    mRNA    linear   EST 22-JAN-1999
DEFINITION  yt73h03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:276316 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   R94003
VERSION     R94003.1  GI:969398
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1  (bases 1 to 391)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1054
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1054   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Homo sapiens"
                     /db_xref="GDB:3799571"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:276316"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       91 a     93 c     85 g    122 t
ORIGIN

  Query Match             39.7%;  Score 27.8;  DB 14;  Length 391;
  Best Local Similarity   65.1%;  Pred. No. 1e+02;
  Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      7 TTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCA 66
          | | | |  | || ||    ||||| ||| ||||  | | |||| |||||    ||||||
Db    338 TACCCTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCA 279

Qy     67 GAT 69
          |||
Db    278 GAT 276


RESULT 8
T97467/c
```

DH10B)"
/note="Vector: pT7T3D-Pac with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared from zebrafish(C32) fin, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is non-normalized. Library was constructed by Ning Wu. NOTE:Clones from this library are only available thru Research Genetics (www.resgen.com)."
BASE COUNT 116 a 187 c 167 g 122 t
ORIGIN

Query Match 42.6%; Score 29.8; DB 13; Length 592;
Best Local Similarity 70.2%; Pred. No. 27;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy      4 CCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
          || |||| || |  ||||| | |    ||||   || ||||||||||| ||| ||||
Db    579 CCTTTCTTGGAATGTGGAAGCGGTTTCTCAGGCTCCCTCTCGGGAAATCGAACCTGA 523
```

RESULT 4
AZ471549/c
LOCUS AZ471549 517 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0286P10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0286P10 F, DNA sequence.
ACCESSION AZ471549
VERSION AZ471549.1 GI:10629674
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 517)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0286 row: P column: 10
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 517.
FEATURES Location/Qualifiers
source 1. .517
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0286P10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 127 a 130 c 133 g 127 t
ORIGIN

Query Match 42.3%; Score 29.6; DB 17; Length 517;
Best Local Similarity 88.9%; Pred. No. 30;
Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      4 CCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCC 39
          |||||||||||||||||||||||||||||    | ||
Db    309 CCCTTCTCGGGAAATGGAAACTGAAAATCTACTTCC 274
```

RESULT 5
BI936879/c
LOCUS BI936879 391 bp mRNA linear EST 18-OCT-2001
DEFINITION dd16b07.y1 Wellcome CRC pSK St 10 5 Xenopus laevis cDNA clone IMAGE:3420612 5' similar to TR:O14597 O14597 NON-FUNCTIONAL FOLATE BINDING PROTEIN. ;, mRNA sequence.
ACCESSION BI936879
VERSION BI936879.1 GI:16251351
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 165.
FEATURES Location/Qualifiers
source 1. .391
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3420612"
/clone_lib="Wellcome CRC pSK St 10 5"
/tissue_type="embryo, stage 10.5"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBluescript SK-; Site_1: NotI; Site_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT 88 a 106 c 90 g 107 t

```
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-CT0249-
            130900-212-b06&t3=2000-09-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 83
            High quality sequence stop: 504.
FEATURES              Location/Qualifiers
     source           1. .506
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="CT0249"
                      /dev_stage="Adult"
                      /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT       154 a    117 c    102 g    133 t
ORIGIN

  Query Match             45.1%;  Score 31.6;  DB 14;  Length 506;
  Best Local Similarity   74.1%;  Pred. No. 7.3;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      14 GAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAG 67
           |||  |||||||||||| ||| ||||||    ||| ||||||  |||||  | ||
Db     220 GAATAGGAAACTGAAATTCAAATCCCTGACTGGGGAATGGATTCTGAATCTTAG 273


RESULT 2
BG142899
LOCUS       BG142899                 540 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  ia90h10.y1 Melton Mouse E16 5 Pancreas Library 2 M16B2 Mus musculus
            cDNA clone IMAGE:5645803 5', mRNA sequence.
ACCESSION   BG142899
VERSION     BG142899.1  GI:12646298
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,
            Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter
            ,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole
            ,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     WashU-Harvard Pancreas EST Project
  JOURNAL   Unpublished (2000)
  COMMENT   Other_ESTs: ia90h10.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            MGI:1852523 This sequence now available from the IMAGE consortium,
            for clone orders contact: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 442.
FEATURES              Location/Qualifiers
     source           1. .540
                      /organism="Mus musculus"
                      /strain="ICR"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5645803"
                      /clone_lib="Melton Mouse E16 5 Pancreas Library 2 M16B2"
                      /sex="Both"
                      /tissue_type="Total pancreas"
                      /dev_stage="Embryonic day 16.5"
                      /lab_host="TOP10"
                      /note="Organ: Pancreas; Vector: pBluescript II SK; Site_1:
                      NotI; Site_2: SalI; Library constructed using SuperScript
                      Plasmid Library kit (Life  Technologies). cDNA made by
                      oligo-dT priming.Size-selected by  column fractionation;
                      average insert size 1.06kb. Primary  library,
                      unamplified."
BASE COUNT       152 a    128 c     99 g    161 t
ORIGIN

  Query Match             42.9%;  Score 30;  DB 12;  Length 540;
  Best Local Similarity   64.3%;  Pred. No. 23;
  Matches   45;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
           | | |||| | || ||||| |||||||||     | | |  ||| |     |||||||
Db      55 GTTTCCTTGTTGGAAAATGAAAACTGAAACCTGAAACTCAACTCTGATTCAGGAAACTTA 114

Qy      61 AAATCAGATC 70
           |||||  | |
Db     115 AAATCGAAAC 124


RESULT 3
BI897688/c
LOCUS       BI897688                 592 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  fn04e10.x1 Zebrafish Research Genetics C32 fin Danio rerio cDNA
            clone 4469466 3' similar to TR:O14597 O14597 NON-FUNCTIONAL FOLATE
            BINDING PROTEIN. ;, mRNA sequence.
ACCESSION   BI897688
VERSION     BI897688.1  GI:16140824
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
  COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by: Steve
            Johnson. DNA Sequencing by: Washington University Genome Sequencing
            Center Clone distribution: IMAGE consortium email:
            info@image.llnl.gov
            Possible reversed clone: polyT not found
            Seq primer: T7 from Gibco
            High quality sequence stop: 469.
FEATURES              Location/Qualifiers
     source           1. .592
                      /organism="Danio rerio"
                      /db_xref="taxon:7955"
                      /clone="4469466"
                      /clone_lib="Zebrafish Research Genetics C32 fin"
                      /tissue_type="Fin"
                      /lab_host="GeneHogs (HS996, a phage-resistant isolate of
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 527.874 Seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-2
Perfect score: 70
Sequence: 1 gatcccttctcgggaaatgg.....tggaaactgaaaatcagatc 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	31.6	45.1	506	14	BQ321577	BQ321577 RC1-CT024
	2	30	42.9	540	12	BG142899	BG142899 ia90h10.y
c	3	29.8	42.6	592	13	BI897688	BI897688 fn04e10.x
c	4	29.6	42.3	517	17	AZ471549	AZ471549 1M0286P10
c	5	28.2	40.3	391	13	BI936879	BI936879 dd16b07.y
c	6	28.2	40.3	540	17	AZ913831	AZ913831 RPCI-24-1
c	7	27.8	39.7	391	14	R94003	R94003 yt73h03.s1
c	8	27.8	39.7	434	14	T97467	T97467 ye53b12.s1
	9	27.6	39.4	502	17	AQ630605	AQ630605 RPCI-11-4
	10	27.6	39.4	656	13	BJ255893	BJ255893 BJ255893
	11	27.6	39.4	658	14	BQ238956	BQ238956 TaE05039B
	12	27.6	39.4	753	14	BQ842141	BQ842141 WHE2989_D
	13	27.4	39.1	279	9	AV096262	AV096262 AV096262
	14	27.4	39.1	451	17	AQ664260	AQ664260 HS_5496_B
c	15	27.4	39.1	747	13	BI938198	BI938198 de36c01.y
	16	27.2	38.9	688	10	AV681534	AV681534 AV681534
	17	27.2	38.9	1037	12	BF979744	BF979744 602288630
c	18	27	38.6	419	14	N59640	N59640 yv74e07.s1
	19	27	38.6	695	17	AZ812128	AZ812128 2M0078B11
	20	27	38.6	719	13	BI855882	BI855882 603383043
	21	26.8	38.3	440	14	BQ398571	BQ398571 NISC_mo09
	22	26.8	38.3	524	13	BJ032573	BJ032573 BJ032573
c	23	26.8	38.3	533	17	AQ623869	AQ623869 HS_5376_A
c	24	26.8	38.3	569	17	AQ695555	AQ695555 HS_2148_A
	25	26.8	38.3	605	10	AV825378	AV825378 AV825378
c	26	26.8	38.3	655	17	BH674985	BH674985 BOMOR25TF
c	27	26.8	38.3	682	17	BH680864	BH680864 BOHTG36TR
c	28	26.8	38.3	710	12	BF119479	BF119479 601758413
	29	26.6	38.0	164	17	AZ745116	AZ745116 RPCI-24-9
c	30	26.6	38.0	243	10	AV355881	AV355881 AV355881
	31	26.6	38.0	331	9	AA728941	AA728941 nw02b09.s
c	32	26.6	38.0	341	12	BG508795	BG508795 sac90d12.
c	33	26.6	38.0	473	17	AZ159253	AZ159253 SP_0062_A
	34	26.6	38.0	550	17	AQ630596	AQ630596 RPCI-11-4
	35	26.6	38.0	575	13	BM383892	BM383892 UI-R-DM1-
c	36	26.6	38.0	1059	13	BM467331	BM467331 AGENCOURT
c	37	26.4	37.7	418	17	AQ210467	AQ210467 HS_2202_B
c	38	26.4	37.7	508	13	BI940410	BI940410 dad74g06.
c	39	26.4	37.7	726	9	AI217912	AI217912 qf49c03.x
c	40	26.2	37.4	225	10	BB337449	BB337449 BB337449
c	41	26.2	37.4	481	14	T56182	T56182 yb79g04.s1
	42	26.2	37.4	534	12	BG447023	BG447023 GA__Eb004
	43	26.2	37.4	603	13	BI503827	BI503827 BB170025B
	44	26.2	37.4	668	13	BJ071703	BJ071703 BJ071703
	45	26	37.1	184	10	AW021778	AW021778 df28b08.y

ALIGNMENTS

RESULT 1
BQ321577
LOCUS BQ321577 506 bp mRNA linear EST 17-MAY-2002
DEFINITION RC1-CT0249-130900-212-b06 CT0249 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ321577
VERSION BQ321577.1 GI:20929928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 506)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Search completed: March 5, 2003, 22:49:56
Job time : 73.3092 secs

```
PS   Claim 25; SEQ ID No 4962; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256


RESULT 14
AAI36372
ID   AAI36372 standard; DNA; 415 BP.
XX
AC   AAI36372;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #5058 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 5058; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256


RESULT 15
AAI04795
ID   AAI04795 standard; DNA; 415 BP.
XX
AC   AAI04795;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #4786 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 4786; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256
```

KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 4890; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

Query Match 36.6%; Score 25.6; DB 22; Length 415;
Best Local Similarity 66.1%; Pred. No. 23;
Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
||||| || ||| ||| |||| || | || || ||| ||| ||||| ||
Db 201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256

RESULT 12
AAK30423
AAK30423 standard; DNA; 415 BP.
XX
AC AAK30423;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 4980.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 4980; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

Query Match 36.6%; Score 25.6; DB 22; Length 415;
Best Local Similarity 66.1%; Pred. No. 23;
Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
||||| || ||| ||| |||| || | || || ||| ||| ||||| ||
Db 201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256

RESULT 13
AAI15029
ID AAI15029 standard; DNA; 415 BP.
XX
AC AAI15029;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4962 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX

```
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256


RESULT 9
ABA56773
ID   ABA56773 standard; DNA; 415 BP.
XX
AC   ABA56773;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #5078.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 1; SEQ ID NO 5078; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256


RESULT 10
ABA26396
ID   ABA26396 standard; DNA; 415 BP.
XX
AC   ABA26396;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #4862 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 1; SEQ ID No 4862; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

  Query Match             36.6%;  Score 25.6;  DB 22;  Length 415;
  Best Local Similarity   66.1%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      15 AAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||  ||   ||| ||| |||| ||  |   || || ||| |||  ||||| ||
Db     201 AAATGACAATAAAAATTCACATCCATTGACCTAAATTGAAAAATGATCATCAGCTC 256


RESULT 11
AAK04899
ID   AAK04899 standard; DNA; 415 BP.
XX
AC   AAK04899;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 4890.
XX
```

CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 551 BP; 187 A; 102 C; 116 G; 146 T; 0 other;

Query Match 38.6%; Score 27; DB 22; Length 551;
Best Local Similarity 66.1%; Pred. No. 8.1;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 11 CGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
Db 193 CTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCAGAT 251

RESULT 7
ABL77594/c
ID ABL77594 standard; cDNA; 481 BP.
XX
AC ABL77594;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:572.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
Claim 1; SEQ ID 572; 489pp; English.

CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 481 BP; 122 A; 83 C; 120 G; 152 T; 4 other;

Query Match 37.4%; Score 26.2; DB 24; Length 481;
Best Local Similarity 66.1%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAA 56
Db 423 GATCCTTCCATGGGAAATGNAAAAAGAAAATGTAACAACATCCCATGCAATTGAAA 368

RESULT 8
ABA46225
ID ABA46225 standard; DNA; 415 BP.
XX
AC ABA46225;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4920.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 1; SEQ ID NO 4920; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 4997; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a

PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 4996; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 551 BP; 187 A; 101 C; 116 G; 147 T; 0 other;

Query Match 38.6%; Score 27; DB 22; Length 551;
Best Local Similarity 66.1%; Pred. No. 8.1;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy      11 CGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           | |  | || ||    ||||| ||| ||||  | | |||| |||||    |||||||||
Db     193 CTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCAGAT 251
```

RESULT 6
AAK91421
ID AAK91421 standard; DNA; 551 BP.
XX
AC AAK91421;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4997.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465567/50.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 901; 562pp; English.
XX
CC The present invention relates to the isolation of novel human colon
CC associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC sequences encoding for them. The sequences of the invention are useful
CC in the diagnosis, treatment, prevention and/or prognosis of disorders
CC of the colon including colon cancer, congenital abnormalities
CC (e.g. atresia and stenosis), bacterial and viral infections,
CC inflammatory bowel disease (IBD), neoplastic cell disorders
CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC The polynucleotides sequences of the invention can also be used in gene
CC therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC novel human colon associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 551 BP; 187 A; 102 C; 116 G; 146 T; 0 other;

Query Match 38.6%; Score 27; DB 22; Length 551;
Best Local Similarity 66.1%; Pred. No. 8.1;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 11 CGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69

Db 193 CTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCAGAT 251

RESULT 5
AAK91420
ID AAK91420 standard; DNA; 551 BP.
XX
AC AAK91420;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4996.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.

```
CC   (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC   colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC   (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC   The polynucleotides sequences of the invention can also be used in gene
CC   therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC   novel human colon associated polypeptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 551 BP; 187 A; 101 C; 116 G; 147 T; 0 other;

  Query Match             38.6%;  Score 27;  DB 22;  Length 551;
  Best Local Similarity   66.1%;  Pred. No. 8.1;
  Matches   39; Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      11 CGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           | |  | || ||    ||||| ||| ||||  | | |||| ||||    |||||||||
Db     193 CTGCTACTGAAACAACAAAATAAGACCCCTGTTTGAGAAAGGGAAATATGAAATCAGAT 251


RESULT 4
AAS40004
     AAS40004 standard; DNA; 551 BP.
XX
AC   AAS40004;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #423 encoding human colon associated polypeptide.
XX
KW   Human; colon cancer; congenital abnormality; infection; colitis;
KW   inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW   intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW   sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW   cytostatic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155302-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01240.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
     18-APR-2000; 2000US-0198123.
     19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
```

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465567/50.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 900; 562pp; English.
XX
CC The present invention relates to the isolation of novel human colon
CC associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC sequences encoding for them. The sequences of the invention are useful
CC in the diagnosis, treatment, prevention and/or prognosis of disorders
CC of the colon including colon cancer, congenital abnormalities
CC (e.g. atresia and stenosis), bacterial and viral infections,
CC inflammatory bowel disease (IBD), neoplastic cell disorders

PT mammalian gene expression, is based on the interferon regulating
PT factor-1 and its binding site -
XX
PS Claim 1; Page 9; 19pp; English.
XX
CC The present invention describes a promoter-transactivator system for
CC inducible high-level expression of mammalian genes, and optionally
CC for control of cell growth. The promoter-transactivator system
CC comprises: (i) a promoter construct (IRFE promoter); and (ii)
CC transactivator construct encoding a fusion protein of IRF-1 (interferon
CC regulating factor-1) and the oestrogen receptor (ER). The IRFE promoter
CC construct has the structure (MPSV-E)-(IRF-1 binding site)-(CMV)-DNA
CC where MPSV-E indicates the myeloproliferative sarcoma virus enhancer
CC repeat given in AAC64591; IRF-1 binding site given in AAC64592; and CMV
CC is the cytomegalovirus minimal promoter given in AAC64593 or their
CC functionally equivalent variants with one or more nucleotides
CC substituted, inserted or deleted. The promoter transactivator system is
CC a transcription regulator. Increased transcription results from binding
CC of the IRFE promoter to the transactivator, which is activated by
CC oestradiol or other ER ligands (these compounds displace the heat-shock
CC protein 90 which normally binds to the IRF-1/ER fusion, preventing its
CC activation). The system is used for increasing production of
CC therapeutically active proteins and where a IRF-1-green fluorescent
CC protein-human ER construct is used, for fluorescence-activated cell
CC sorting/analysis of transformed cells.
XX
SQ Sequence 70 BP; 25 A; 15 C; 15 G; 15 T; 0 other;

Query Match 100.0%; Score 70; DB 21; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60

Qy      61 AAATCAGATC 70
           ||||||||||
Db      61 AAATCAGATC 70
```

RESULT 2
AAS80058
ID AAS80058 standard; cDNA; 2286 BP.
XX
AC AAS80058;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15862.
XX
Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG15871.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 15862; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2286 BP; 536 A; 705 C; 650 G; 395 T; 0 other;

Query Match 39.4%; Score 27.6; DB 23; Length 2286;
Best Local Similarity 67.2%; Pred. No. 7.1;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```
Qy      13 GGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATC 70
           |||||||||||  ||||   ||| || |  |   |   ||||||| |||| | |||||
Db    1705 GGAAATGGAAATGGAAAGCAAGACCCATATTAATGTCTTGGAAACAGAAAGCAAGATC 1762
```

RESULT 3
AAS40003
ID AAS40003 standard; DNA; 551 BP.
XX
AC AAS40003;
XX
DT 17-DEC-2001 (first entry)
XX
DE Genomic sequence #422 encoding human colon associated polypeptide.
XX
KW Human; colon cancer; congenital abnormality; infection; colitis;
KW inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW cytostatic; ds.
XX
OS Homo sapiens.
XX
PN WO200155302-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01240.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:12 ; Search time 68.3092 Seconds
(without alignments)
2307.741 Million cell updates/sec

Title: US-09-980-548-2
Perfect score: 70
Sequence: 1 gatcccttctcgggaaatgg.....tggaaactgaaaatcagatc 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	70	100.0	70	21	AAC64592	Firefly IRF-1 bind
	2	27.6	39.4	2286	23	AAS80058	DNA encoding novel
	3	27	38.6	551	22	AAS40003	Genomic sequence #
	4	27	38.6	551	22	AAS40004	Genomic sequence #
	5	27	38.6	551	22	AAK91420	Human digestive sy
	6	27	38.6	551	22	AAK91421	Human digestive sy
c	7	26.2	37.4	481	24	ABL77594	Human ovarian canc
	8	25.6	36.6	415	22	ABA46225	Human breast cell
	9	25.6	36.6	415	22	ABA56773	Human foetal liver
	10	25.6	36.6	415	22	ABA26396	Probe #4862 for ge
	11	25.6	36.6	415	22	AAK04899	Human brain expres
	12	25.6	36.6	415	22	AAK30423	Human bone marrow
	13	25.6	36.6	415	22	AAI15029	Probe #4962 for ge
	14	25.6	36.6	415	22	AAI36372	Probe #5058 used t
	15	25.6	36.6	415	22	AAI04795	Probe #4786 used t
	16	25.6	36.6	415	24	ABS05034	Human genome-deriv
c	17	25.6	36.6	8981	21	AAC69138	Human ABC1 gene ex
c	18	25.6	36.6	183999	22	AAF92831	Human ABC1 genomic
c	19	25.4	36.3	715	20	AAX37514	Human secreted pro
	20	25.4	36.3	90104	23	ABL12402	Drosophila melanog
	21	25.4	36.3	684707	24	ABQ67196	Listeria innocua c
c	22	25.4	36.3	3011208	24	ABQ69245	Listeria innocua D
c	23	25.2	36.0	348	22	AAS02035	Bladder cancer-ass
	24	25.2	36.0	1230025	20	AAX91990	Nucleotide sequenc
	25	25	35.7	976	23	ABV04483	Human prostate exp
	26	24.8	35.4	310	20	AAV89187	EST clone CH699.
c	27	24.8	35.4	685	21	AAC53345	Arabidopsis thalia
	28	24.8	35.4	705	22	AAF22680	Human gastric canc
	29	24.8	35.4	887	22	AAF22682	Human gastric canc
	30	24.8	35.4	996	21	AAA05586	Streptococcus pneu
	31	24.8	35.4	1789	20	AAX40059	Colon cancer assoc
	32	24.8	35.4	2353	19	AAV40542	Homo sapiens secre
	33	24.8	35.4	2566	22	AAF59641	Human cell cycle a
c	34	24.8	35.4	7568	23	ABL14768	Drosophila melanog
c	35	24.8	35.4	11887	19	AAV52279	Streptococcus pneu
c	36	24.8	35.4	49312	21	AAH51594	Human genomic sequ
	37	24.8	35.4	140036	24	AAS98600	Human genomic DNA
c	38	24.6	35.1	1510	22	AAH99084	Human EST-derived
c	39	24.6	35.1	3884	22	AAH98627	Human EST-derived
c	40	24.6	35.1	5397	23	AAS66000	DNA encoding novel
	41	24.6	35.1	9909	19	AAV52145	Streptococcus pneu
	42	24.6	35.1	2365589	24	ABA90521	Genomic sequence o
	43	24.4	34.9	244	22	AAS23933	Human ovarian PCR-
	44	24.4	34.9	244	22	AAH82491	Human ovarian tumo
c	45	24.4	34.9	1555	21	AAC41543	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAC64592
ID AAC64592 standard; DNA; 70 BP.
XX
AC AAC64592;
XX
DT 15-FEB-2001 (first entry)
XX
DE Firefly IRF-1 binding site nucleotide sequence SEQ ID NO:2.
XX
KW Myeloproliferative sarcoma virus; MPSV; MPSV-E; enhancer; CMV;
KW cytomegalovirus; firefly; IRF-1 binding site; minimal promoter;
KW interferon regulatory factor 1 binding site; transcription regulator;
KW promoter transactivator system; ds.
XX
OS Photinus pyralis.
XX
PN EP1046710-A1.
XX
PD 25-OCT-2000.
XX
PF 23-APR-1999; 99EP-0108068.
XX
PR 23-APR-1999; 99EP-0108068.
XX
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI Mueller PP, Geserick C, Schroeder K, Hauser H;
XX
DR WPI; 2000-648930/63.
XX
PT Promoter-transactivator system, useful for inducing high level

```
misc_feature    5440. .5449
                /note="match to EST BF354907 (NID:g11313981)"
repeat_region   7059. .7366
                /rpt_family="Alu"
repeat_region   7340. .7386
                /rpt_family="(A)n"
repeat_region   7392. .7550
                /rpt_family="L1"
misc_feature    7719. .8230
                /note="match to EST AI459058 (NID:g4311637) tk03e05.x1"
repeat_region   8265. .8388
                /rpt_family="L2"
repeat_region   9415. .9466
                /rpt_family="(TA)n"
repeat_region   10207. .10258
                /rpt_family="(TTTA)n"
repeat_region   10232. .10370
                /rpt_family="Alu"
repeat_region   11214. .11365
                /rpt_family="MIR"
repeat_region   11400. .11620
                /rpt_family="MIR"
repeat_region   11641. .11932
                /rpt_family="Alu"
repeat_region   12012. .12105
                /rpt_family="MIR"
repeat_region   12631. .12754
                /rpt_family="MIR"
repeat_region   13218. .13336
                /rpt_family="MIR"
repeat_region   13328. .13569
                /rpt_family="L2"
repeat_region   13560. .13648
                /rpt_family="MIR"
repeat_region   13688. .13821
                /rpt_family="MIR"
repeat_region   13907. .14441
                /rpt_family="L2"
repeat_region   14542. .14580
                /rpt_family="L2"
misc_feature    15070. .15475
                /note="match to EST AI032114 (NID:g3250326) os76a07.s1"
repeat_region   15239. .15525
                /rpt_family="Alu"
misc_feature    15610. .15611
                /note="match to EST AI032114 (NID:g3250326) os76a07.s1"
repeat_region   16114. .16234
                /rpt_family="L1"
repeat_region   16517. .16784
                /rpt_family="L1"
repeat_region   16787. .16863
                /rpt_family="MER2_type"
repeat_region   16844. .16907
                /rpt_family="AT_rich"
repeat_region   16886. .17188
                /rpt_family="Alu"
repeat_region   17246. .17352
                /rpt_family="MER2_type"
repeat_region   17353. .17643
                /rpt_family="L1"
repeat_region   17657. .18139
                /rpt_family="L1"
repeat_region   18140. .18990
                /rpt_family="L1"
repeat_region   18617. .18665
                /rpt_family="AT_rich"
repeat_region   19048. .19330
                /rpt_family="L1"
repeat_region   19333. .20537
                /rpt_family="MER2_type"
repeat_region   20539. .20717
                /rpt_family="Alu"
repeat_region   20718. .21877
                /rpt_family="MER2_type"
repeat_region   21880. .22318
                /rpt_family="L1"
repeat_region   22315. .22543
                /rpt_family="L1"
repeat_region   22572. .22606
                /rpt_family="Mariner"
repeat_region   22616. .22975
                /rpt_family="L1"
repeat_region   23377. .23609
                /rpt_family="MIR"
repeat_region   24093. .24207
                /rpt_family="L1"
repeat_region   24445. .24741
                /rpt_family="Alu"
repeat_region   24719. .24741
                /rpt_family="(A)n"
misc_feature    25224. .25290
                /note="match to EST AL561641 (NID:g12909272)"
misc_feature    25225. .25641
                /note="match to EST BG192864 (NID:g13714551)"
misc_feature    25230. .25290
                /note="match to EST BE561346 (NID:g9805053)"
misc_feature    25230. .25290

  Query Match             41.4%;  Score 29;  DB 9;  Length 165373;
  Best Local Similarity   63.8%;  Pred. No. 41;
  Matches   44;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
           || ||  ||||   |||   |||   | ||  | ||| ||||| |||||  |||  | |
Db  113685 GACCCTATCTCTTAAAAAAAAAAAACAGAAAAATATCACTTCTGGGGAAACAGAACTTAA 113744

Qy      61 AAATCAGAT 69
           | | |||||
Db  113745 AGAACAGAT 113753


Search completed: March  5, 2003, 23:34:38
Job time : 766.676 secs
```

```
                     /note="assembly_fragment"
     misc_feature    2879. .6107
                     /note="assembly_fragment"
     misc_feature    6208. .7623
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    7724. .11841
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    11942. .18456
                     /note="assembly_fragment"
     misc_feature    18557. .27382
                     /note="assembly_fragment"
     misc_feature    27483. .37124
                     /note="assembly_fragment"
     misc_feature    37225. .47276
                     /note="assembly_fragment"
     misc_feature    47377. .66713
                     /note="assembly_fragment"
     misc_feature    66814. .89140
                     /note="assembly_fragment"
     misc_feature    89241. .116360
                     /note="assembly_fragment"
     misc_feature    116461. .141816
                     /note="assembly_fragment"
     misc_feature    141917. .176775
                     /note="assembly_fragment"
BASE COUNT    42683 a  45397 c  44545 g  42948 t   1202 others
ORIGIN

  Query Match             41.7%;  Score 29.2;  DB 2;  Length 176775;
  Best Local Similarity   74.0%;  Pred. No. 35;
  Matches   37;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      13 GGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAA 62
           |||| |||| || | || ||||| | |||| | | ||| |||||  ||||
Db 104829 GGAATTGGAGACAGCAAGTCAGAACTCTTCCCTGTAAAGGGAAAGAGAAA 104780


RESULT 15
AC067957
LOCUS       AC067957               165373 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-82K13 from 2, complete sequence.
ACCESSION   AC067957
VERSION     AC067957.7  GI:15638744
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    ERENCE  1  (bases 1 to 165373)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 165373)
  AUTHORS   Isak,A., Kozlowicz,A., Dixon,R. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-82K13
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 165373)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 165373)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 165373)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:14670148.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0082K13
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-111K7; the clone sequenced
            to the right is RP11-89K21, 2000 bp overlap.  Actual start of this
            clone is at base position 1 of RP11-82K13; actual end is at base
            position 29707 of RP11-89K21.

            Data from AC012455, AC067791, and AC062030 was used to finish this
            clone, AC067957.

            Polymorphisms exist between AC067957 and AC067791.
FEATURES                Location/Qualifiers
     source             1. .165373
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-82K13"
                        /clone_lib="RPCI-11"
     misc_feature       3097. .3427
                        /note="match to EST H64341 (NID:g1023081) yr71a12.r1"
     repeat_region      4326. .4440
                        /rpt_family="MIR"
     repeat_region      4486. .4799
                        /rpt_family="Alu"
     repeat_region      4770. .4820
                        /rpt_family="(A)n"
     misc_feature       5390. .5549
                        /note="match to EST AA485569 (NID:g2214788) ab09g06.r1"
```

```
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-149P14 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-149P14 is at 1 in this sequence.
            The true left end of clone RP4-539L13 is at 108122 in this
            sequence. The true right end of clone RP11-192B7 is at 58622 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .110121
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-149P14"
                     /clone_lib="RPCI-11.1"
     misc_feature    6650. .6686
                     /note="Sequence from overlapping clone RP11-192B7
                     (AL592314). Assembly confirmed by restriction digest."
     misc_feature    21058. .21123
                     /note="Sequence from overlapping clone RP11-192B7
                     (AL592314). Assembly confirmed by restriction digest."
     misc_feature    50716. .50718
                     /note="Sequence from overlapping clone RP11-192B7
                     (AL592314). Assembly confirmed by restriction digest."
     misc_feature    54763. .54871
                     /note="Sequence from overlapping clone RP11-192B7
                     (AL592314). Assembly confirmed by restriction digest."
     misc_feature    86404. .86414
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    86780. .86903
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
BASE COUNT    25978 a  28405 c  27853 g  27885 t
ORIGIN

  Query Match             41.7%;  Score 29.2;  DB 9;  Length 110121;
  Best Local Similarity   74.0%;  Pred. No. 35;
  Matches   37;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      13 GGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAA 62
           |||| |||| || | || ||||| | |||| | | ||| |||||  ||||
Db   26583 GGAATTGGAGACAGCAAGTCAGAACTCTTCCCTGTAAAGGGAAAGAGAAA 26534


RESULT 14
AC016978/c
LOCUS       AC016978               176775 bp    DNA     linear   HTG 09-SEP-2000
DEFINITION  Homo sapiens clone RP11-5N7, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC016978
VERSION     AC016978.3  GI:10045327
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176775)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-5N7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176775)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,
            Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,
            Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,
            Karatas,A., Klein,J., Landers,T., Lehoczky,J., Lieu,C., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,
            Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 9, 2000 this sequence version replaced gi:6648261.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2875
                Center clone name: 5_N_7
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 166773 bases at least Q40
                Consensus quality: 172242 bases at least Q30
                Consensus quality: 174181 bases at least Q20
                Insert size: 85000; agarose-fp
                Insert size: 175575; sum-of-contigs
                Quality coverage: 8.8 in Q20 bases; agarose-fp
                Quality coverage: 4.3 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2778: contig of 2778 bp in length
            *     2779  2878: gap of      100 bp
            *     2879     6107: contig of 3229 bp in length
            *     6108  6207: gap of      100 bp
            *     6208     7623: contig of 1416 bp in length
            *     7624  7723: gap of      100 bp
            *     7724    11841: contig of 4118 bp in length
            *    11842 11941: gap of      100 bp
            *    11942    18456: contig of 6515 bp in length
            *    18457 18556: gap of      100 bp
            *    18557    27382: contig of 8826 bp in length
            *    27383 27482: gap of      100 bp
            *    27483    37124: contig of 9642 bp in length
            *    37125 37224: gap of      100 bp
            *    37225    47276: contig of 10052 bp in length
            *    47277 47376: gap of      100 bp
            *    47377    66713: contig of 19337 bp in length
            *    66714 66813: gap of      100 bp
            *    66814    89140: contig of 22327 bp in length
            *    89141 89240: gap of      100 bp
            *    89241   116360: contig of 27120 bp in length
            *   116361 116460: gap of      100 bp
            *   116461   141816: contig of 25356 bp in length
            *   141817 141916: gap of      100 bp
            *   141917   176775: contig of 34859 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176775
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-5N7"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2778
```

```
*    25929    26799: contig of 871 bp in length
*    26800 26899: gap of      100 bp
*    26900    27808: contig of 909 bp in length
*    27809 27908: gap of      100 bp
*    27909    28812: contig of 904 bp in length
*    28813 28912: gap of      100 bp
*    28913    29813: contig of 901 bp in length
*    29814 29913: gap of      100 bp
*    29914    30807: contig of 894 bp in length
*    30808 30907: gap of      100 bp
*    30908    31790: contig of 883 bp in length
*    31791 31890: gap of      100 bp
*    31891    32777: contig of 887 bp in length
*    32778 32877: gap of      100 bp
*    32878    33789: contig of 912 bp in length
*    33790 33889: gap of      100 bp
*    33890    34792: contig of 903 bp in length
*    34793 34892: gap of      100 bp
*    34893    35795: contig of 903 bp in length
*    35796 35895: gap of      100 bp
*    35896    36798: contig of 903 bp in length
*    36799 36898: gap of      100 bp
*    36899    37791: contig of 893 bp in length
*    37792 37891: gap of      100 bp
*    37892    38732: contig of 841 bp in length
*    38733 38832: gap of      100 bp
*    38833    39746: contig of 914 bp in length
*    39747 39846: gap of      100 bp
*    39847    40745: contig of 899 bp in length
*    40746 40845: gap of      100 bp
*    40846    41743: contig of 898 bp in length
*    41744 41843: gap of      100 bp
*    41844    42723: contig of 880 bp in length
*    42724 42823: gap of      100 bp
*    42824    43687: contig of 864 bp in length
*    43688 43787: gap of      100 bp
*    43788    44683: contig of 896 bp in length
*    44684 44783: gap of      100 bp
*    44784    45685: contig of 902 bp in length
*    45686 45785: gap of      100 bp
*    45786    46689: contig of 904 bp in length
*    46690 46789: gap of      100 bp
*    46790    47690: contig of 901 bp in length
*    47691 47790: gap of      100 bp
*    47791    48690: contig of 900 bp in length
*    48691 48790: gap of      100 bp
*    48791    49709: contig of 919 bp in length
*    49710 49809: gap of      100 bp
*    49810    50684: contig of 875 bp in length
*    50685 50784: gap of      100 bp
*    50785    51684: contig of 900 bp in length
*    51685 51784: gap of      100 bp
*    51785    52685: contig of 901 bp in length
*    52686 52785: gap of      100 bp
*    52786    53691: contig of 906 bp in length
*    53692 53791: gap of      100 bp
*    53792    54688: contig of 897 bp in length
*    54689 54788: gap of      100 bp
*    54789    55705: contig of 917 bp in length
*    55706 55805: gap of      100 bp
*    55806    56714: contig of 909 bp in length
*    56715 56814: gap of      100 bp
*    56815    57704: contig of 890 bp in length
*    57705 57804: gap of      100 bp
*    57805    58702: contig of 898 bp in length
*    58703 58802: gap of      100 bp
*    58803    59683: contig of 881 bp in length
*    59684 59783: gap of      100 bp
*    59784    60662: contig of 879 bp in length
*    60663 60762: gap of      100 bp
*    60763    61642: contig of 880 bp in length
*    61643 61742: gap of      100 bp
*    61743    62617: contig of 875 bp in length
*    62618 62717: gap of      100 bp
*    62718    63623: contig of 906 bp in length
*    63624 63723: gap of      100 bp
*    63724    64617: contig of 894 bp in length
*    64618 64717: gap of      100 bp
*    64718    65635: contig of 918 bp in length
*    65636 65735: gap of      100 bp
*    65736    66612: contig of 877 bp in length
*    66613 66712: gap of      100 bp
*    66713    67591: contig of 879 bp in length
*    67592 67691: gap of      100 bp
*    67692    68598: contig of 907 bp in length
*    68599 68698: gap of      100 bp
*    68699    69594: contig of 896 bp in length
*    69595 69694: gap of      100 bp
*    69695    70501: contig of 807 bp in length
*    70502 70601: gap of      100 bp
*    70602    71452: contig of 851 bp in length
*    71453 71552: gap of      100 bp
*    71553    72370: contig of 818 bp in length

  Query Match             42.3%;  Score 29.6;  DB 2;  Length 90747;
  Best Local Similarity   88.9%;  Pred. No. 26;
  Matches   32;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 CCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCC 39
           ||||||||||||||||||||||||||||||   | ||
Db    1466 CCCTTCTCGGGAAATGGAAACTGAAAATCTACTTCC 1431


RESULT 13
AL590989/c
LOCUS       AL590989              110121 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-149P14 on chromosome 1, complete
            sequence.
ACCESSION   AL590989
VERSION     AL590989.15  GI:16973092
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110121)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:16508264.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-149P14 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-19L12 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6.

FEATURES Location/Qualifiers
source 1. .197743
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-19L12"
/clone_lib="RPCI-23"

BASE COUNT 52001 a 46653 c 47240 g 51849 t
ORIGIN

Query Match 42.9%; Score 30; DB 10; Length 197743;
Best Local Similarity 64.3%; Pred. No. 19;
Matches 45; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
           | | |||| | || ||||| |||||||||     | | |  ||| |    |||||||
Db 131238 GTTTCCTTGTTGGAAAATGAAAACTGAAACCTGAAACTCAACTCTGATTCAGGAAACTTA 131297

Qy      61 AAATCAGATC 70
           |||||  | |
Db 131298 AAATCGAAAC 131307
```

RESULT 12
AC011975/c
LOCUS AC011975 90747 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-15M4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC011975
VERSION AC011975.4 GI:9123856
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 90747)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-15M4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 90747)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 13, 2000 this sequence version replaced gi:7144916.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L3491
Center clone name: 15_M_4
----------.
* NOTE: This record contains 93 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 909: contig of 909 bp in length
* 910 1009: gap of 100 bp
* 1010 1932: contig of 923 bp in length
* 1933 2032: gap of 100 bp
* 2033 2935: contig of 903 bp in length
* 2936 3035: gap of 100 bp
* 3036 3899: contig of 864 bp in length
* 3900 3999: gap of 100 bp
* 4000 4936: contig of 937 bp in length
* 4937 5036: gap of 100 bp
* 5037 5929: contig of 893 bp in length
* 5930 6029: gap of 100 bp
* 6030 6914: contig of 885 bp in length
* 6915 7014: gap of 100 bp
* 7015 7879: contig of 865 bp in length
* 7880 7979: gap of 100 bp
* 7980 8857: contig of 878 bp in length
* 8858 8957: gap of 100 bp
* 8958 9824: contig of 867 bp in length
* 9825 9924: gap of 100 bp
* 9925 10809: contig of 885 bp in length
* 10810 10909: gap of 100 bp
* 10910 11791: contig of 882 bp in length
* 11792 11891: gap of 100 bp
* 11892 12803: contig of 912 bp in length
* 12804 12903: gap of 100 bp
* 12904 13787: contig of 884 bp in length
* 13788 13887: gap of 100 bp
* 13888 14764: contig of 877 bp in length
* 14765 14864: gap of 100 bp
* 14865 15761: contig of 897 bp in length
* 15762 15861: gap of 100 bp
* 15862 16756: contig of 895 bp in length
* 16757 16856: gap of 100 bp
* 16857 17757: contig of 901 bp in length
* 17758 17857: gap of 100 bp
* 17858 18757: contig of 900 bp in length
* 18758 18857: gap of 100 bp
* 18858 19751: contig of 894 bp in length
* 19752 19851: gap of 100 bp
* 19852 20764: contig of 913 bp in length
* 20765 20864: gap of 100 bp
* 20865 21794: contig of 930 bp in length
* 21795 21894: gap of 100 bp
* 21895 22797: contig of 903 bp in length
* 22798 22897: gap of 100 bp
* 22898 23808: contig of 911 bp in length
* 23809 23908: gap of 100 bp
* 23909 24843: contig of 935 bp in length
* 24844 24943: gap of 100 bp
* 24944 25828: contig of 885 bp in length
* 25829 25928: gap of 100 bp

```
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          ------------- Project Information
            Center project name: L5392
            Center clone name: 127_H_5
          -------------
FEATURES             Location/Qualifiers
     source          1. .160883
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-127H5"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   2879. .3178
                     /rpt_family="AluSc"
     repeat_region   3814. .3842
                     /rpt_family="AT_rich"
     repeat_region   complement(4030. .4462)
                     /rpt_family="L2"
     repeat_region   4486. .4540
                     /rpt_family="(TTCA)n"
     repeat_region   complement(4550. .4830)
                     /rpt_family="L1MC1"
     repeat_region   4831. .4920
                     /rpt_family="(CTATA)n"
     repeat_region   complement(4927. .5161)
                     /rpt_family="L1MC1"
     repeat_region   complement(6250. .6361)
                     /rpt_family="MIR"
     repeat_region   6362. .6557
                     /rpt_family="MER63A"
     repeat_region   complement(6558. .6698)
                     /rpt_family="MIR"
     repeat_region   complement(6712. .6865)
                     /rpt_family="MIR3"
     repeat_region   6901. .6927
                     /rpt_family="(T)n"
     repeat_region   complement(8668. .9158)
                     /rpt_family="LTR33"
     repeat_region   complement(9223. .9528)
                     /rpt_family="AluJb"
     repeat_region   9962. .10210
                     /rpt_family="AluSx"
     repeat_region   12393. .12953
                     /rpt_family="L2"
     repeat_region   complement(13061. .13176)
                     /rpt_family="FLAM_A"
     repeat_region   complement(13713. .13976)
                     /rpt_family="LTR16A1"
     repeat_region   14722. .14907
                     /rpt_family="MIR"
     repeat_region   14923. .15077
                     /rpt_family="MIR"
     repeat_region   15547. .15742
                     /rpt_family="MIR"
     repeat_region   complement(15987. .16070)
                     /rpt_family="L2"
     repeat_region   complement(16162. .16216)
                     /rpt_family="MER103"
     repeat_region   16855. .16901
                     /rpt_family="L2"
     repeat_region   17188. .17240
                     /rpt_family="(TG)n"
     repeat_region   complement(17279. .17346)
                     /rpt_family="L2"
     repeat_region   complement(17721. .18015)
                     /rpt_family="AluY"
     repeat_region   complement(18026. .18581)
                     /rpt_family="L1ME1"
     repeat_region   complement(18749. .18863)
                     /rpt_family="FLAM_A"
     repeat_region   18961. .19194
                     /rpt_family="MIR"
     repeat_region   complement(19347. .19420)
                     /rpt_family="L1PA6"
     repeat_region   19461. .19550
                     /rpt_family="L2"
     repeat_region   19562. .19591
                     /rpt_family="(TGAA)n"
     repeat_region   complement(20511. .20745)
                     /rpt_family="MIR"
     repeat_region   complement(20999. .21095)
                     /rpt_family="L2"
     repeat_region   21284. .21520
                     /rpt_family="MLT1J1"
     repeat_region   21949. .21991
                     /rpt_family="(TTTTG)n"
     repeat_region   23647. .23713
                     /rpt_family="(CA)n"
     repeat_region   24203. .24340
                     /rpt_family="MIR"
     repeat_region   complement(24837. .25022)
                     /rpt_family="MIR3"
     repeat_region   complement(25220. .25371)
                     /rpt_family="L2"
     repeat_region   25505. .25811
                     /rpt_family="ERVL-F"
     repeat_region   25815. .26206
                     /rpt_family="MLT2F"
     repeat_region   26278. .26373
                     /rpt_family="MLT2F"
     repeat_region   26395. .26663
                     /rpt_family="L1PA16"
     repeat_region   26664. .26953
                     /rpt_family="AluSx"
     repeat_region   26954. .27438

  Query Match             42.9%;  Score 30;  DB 9;  Length 160883;
  Best Local Similarity   72.2%;  Pred. No. 19;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       3 TCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAA 56
           |||| |||| ||||||||  ||||||| | ||  ||| ||| || |  ||  ||
Db   42450 TCCCGTCTCAGGAAATGGCTACTGAAATTTAGGACCCCTCTTGGCAGTTGTTAA 42397


RESULT 11
AL591711
LOCUS       AL591711              197743 bp    DNA     linear   ROD 25-MAY-2002
DEFINITION  Mouse DNA sequence from clone RP23-19L12 on chromosome 2, complete
            sequence.
ACCESSION   AL591711
VERSION     AL591711.22  GI:21217770
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197743)
  AUTHORS   Johnson,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 27, 2002 this sequence version replaced gi:21068504.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

```
            Center: Keio University, School of Medicine
            Center code: Keio
            Web site: http://www.dmb.med.keio.ac.jp/
            Contact: nshimizu@dmb.med.keio.ac.jp
          -------------- Project Information
            Center clone name: KB1568A9
          -------------- Summary Statistics
            Sequencing vector: pUC118; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 147844 bases at least Q40
            Consensus quality: 148297 bases at least Q30
            Consensus quality: 148363 bases at least Q20
            Insert size: 148555; sum-of-contigs
            Quality coverage: 8.0x in Q20 bases; sum-of-contigs.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    46916: contig of 46916 bp in length
          *   46917 47016: gap of      100 bp
          *   47017   148555: contig of 101539 bp in length.
FEATURES             Location/Qualifiers
     source          1. .148555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1568A9"
                     /cell_line="FLEB14-14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     misc_feature    1. .46916
                     /note="assembly_fragment clone_end:M13 vector_side: right"
     misc_feature    47017. .148555
                     /note="assembly_fragment clone_end:M13RV vector_side:
                     left"
BASE COUNT    43150 a  26939 c  28800 g  49559 t    107 others
ORIGIN

  Query Match             42.9%;  Score 30;  DB 2;  Length 148555;
  Best Local Similarity   72.2%;  Pred. No. 19;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       3 TCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAA 56
           |||| |||| |||||||| |||||||  | ||  ||| ||| || |  ||  ||
Db   82868 TCCCGTCTCAGGAAATGGCTACTGAAATTTAGGACCCCTCTTGGCAGTTGTTAA 82815


RESULT 10
AC021498/c
LOCUS       AC021498              160883 bp    DNA     linear   PRI 20-FEB-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-127H5, complete sequence.
ACCESSION   AC021498
VERSION     AC021498.12  GI:18767547
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-127H5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 160883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 160883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 20, 2002 this sequence version replaced gi:18677531.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
```

```
*    12844    13601: contig of 758 bp in length
*    13602 13701: gap of      100 bp
*    13702    14466: contig of 765 bp in length
*    14467 14566: gap of      100 bp
*    14567    15316: contig of 750 bp in length
*    15317 15416: gap of      100 bp
*    15417    16176: contig of 760 bp in length
*    16177 16276: gap of      100 bp
*    16277    17027: contig of 751 bp in length
*    17028 17127: gap of      100 bp
*    17128    17882: contig of 755 bp in length
*    17883 17982: gap of      100 bp
*    17983    18742: contig of 760 bp in length
*    18743 18842: gap of      100 bp
*    18843    19613: contig of 771 bp in length
*    19614 19713: gap of      100 bp
*    19714    20504: contig of 791 bp in length
*    20505 20604: gap of      100 bp
*    20605    21387: contig of 783 bp in length
*    21388 21487: gap of      100 bp
*    21488    22228: contig of 741 bp in length
*    22229 22328: gap of      100 bp
*    22329    23101: contig of 773 bp in length
*    23102 23201: gap of      100 bp
*    23202    23946: contig of 745 bp in length
*    23947 24046: gap of      100 bp
*    24047    24817: contig of 771 bp in length
*    24818 24917: gap of      100 bp
*    24918    25688: contig of 771 bp in length
*    25689 25788: gap of      100 bp
*    25789    26552: contig of 764 bp in length
*    26553 26652: gap of      100 bp
*    26653    27392: contig of 740 bp in length
*    27393 27492: gap of      100 bp
*    27493    28254: contig of 762 bp in length
*    28255 28354: gap of      100 bp
*    28355    29123: contig of 769 bp in length
*    29124 29223: gap of      100 bp
*    29224    29987: contig of 764 bp in length
*    29988 30087: gap of      100 bp
*    30088    30857: contig of 770 bp in length
*    30858 30957: gap of      100 bp
*    30958    31731: contig of 774 bp in length
*    31732 31831: gap of      100 bp
*    31832    32607: contig of 776 bp in length
*    32608 32707: gap of      100 bp
*    32708    33473: contig of 766 bp in length
*    33474 33573: gap of      100 bp
*    33574    34334: contig of 761 bp in length
*    34335 34434: gap of      100 bp
*    34435    35181: contig of 747 bp in length
*    35182 35281: gap of      100 bp
*    35282    36042: contig of 761 bp in length
*    36043 36142: gap of      100 bp
*    36143    36895: contig of 753 bp in length
*    36896 36995: gap of      100 bp
*    36996    37719: contig of 724 bp in length
*    37720 37819: gap of      100 bp
*    37820    38573: contig of 754 bp in length
*    38574 38673: gap of      100 bp
*    38674    39435: contig of 762 bp in length
*    39436 39535: gap of      100 bp
*    39536    40309: contig of 774 bp in length
*    40310 40409: gap of      100 bp
*    40410    41198: contig of 789 bp in length
*    41199 41298: gap of      100 bp
*    41299    42068: contig of 770 bp in length
*    42069 42168: gap of      100 bp
*    42169    42874: contig of 706 bp in length
*    42875 42974: gap of      100 bp
*    42975    43743: contig of 769 bp in length
*    43744 43843: gap of      100 bp
*    43844    44600: contig of 757 bp in length
*    44601 44700: gap of      100 bp
*    44701    45463: contig of 763 bp in length
*    45464 45563: gap of      100 bp
*    45564    46328: contig of 765 bp in length
*    46329 46428: gap of      100 bp
*    46429    47192: contig of 764 bp in length
*    47193 47292: gap of      100 bp
*    47293    48058: contig of 766 bp in length
*    48059 48158: gap of      100 bp
*    48159    48902: contig of 744 bp in length
*    48903 49002: gap of      100 bp
*    49003    49782: contig of 780 bp in length
*    49783 49882: gap of      100 bp
*    49883    50635: contig of 753 bp in length
*    50636 50735: gap of      100 bp
*    50736    51483: contig of 748 bp in length
*    51484 51583: gap of      100 bp
*    51584    52343: contig of 760 bp in length
*    52344 52443: gap of      100 bp
*    52444    53244: contig of 801 bp in length
*    53245 53344: gap of      100 bp
*    53345    54115: contig of 771 bp in length
*    54116 54215: gap of      100 bp
*    54216    54984: contig of 769 bp in length
*    54985 55084: gap of      100 bp
*    55085    55863: contig of 779 bp in length
*    55864 55963: gap of      100 bp
*    55964    56739: contig of 776 bp in length
*    56740 56839: gap of      100 bp
*    56840    57606: contig of 767 bp in length
*    57607 57706: gap of      100 bp
*    57707    58481: contig of 775 bp in length
*    58482 58581: gap of      100 bp
*    58582    59341: contig of 760 bp in length
*    59342 59441: gap of      100 bp
*    59442    60174: contig of 733 bp in length
*    60175 60274: gap of      100 bp
*    60275    61039: contig of 765 bp in length

  Query Match          42.9%;  Score 30;  DB 2;  Length 62795;
  Best Local Similarity  72.2%;  Pred. No. 19;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       3 TCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAA 56
           |||| |||| ||||||||  |||||||  | ||  ||| ||| || |  ||  ||
Db   31977 TCCCGTCTCAGGAAATGGCTACTGAAATTTAGGACCCCTCTTGGCAGTTGTTAA 31924


RESULT 9
AP004711/c
LOCUS       AP004711              148555 bp    DNA     linear   HTG 02-FEB-2002
DEFINITION  Homo sapiens chromosome 8 clone KB1568A9 map 8q23, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AP004711
VERSION     AP004711.1  GI:18479142
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN; HTGS_DRAFT.
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio
            BAC library clone:KB1568A9.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Homo sapiens DNA chromosome 8 DRAFT SEQUENCE
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 148555)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     ---------- Genome Center
```

```
          *   171141   207659: contig of 36519 bp in length
          *   207660 207759: gap of      100 bp
          *   207760   270196: contig of 62437 bp in length
          *   270197 270296: gap of      100 bp
          *   270297   270795: contig of 499 bp in length.
FEATURES             Location/Qualifiers
     source          1. .270795
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-39M18"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .51876
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    51977. .52976
                     /note="assembly_fragment"
     misc_feature    53077. .54220
                     /note="assembly_fragment"
     misc_feature    54321. .55540
                     /note="assembly_fragment"
     misc_feature    55641. .56917
                     /note="assembly_fragment"
     misc_feature    57018. .58094
                     /note="assembly_fragment"
     misc_feature    58195. .130154
                     /note="assembly_fragment"
     misc_feature    130255. .132870
                     /note="assembly_fragment"
     misc_feature    132971. .143777
                     /note="assembly_fragment"
     misc_feature    143878. .156806
                     /note="assembly_fragment"
     misc_feature    156907. .171040
                     /note="assembly_fragment"
     misc_feature    171141. .207659
                     /note="assembly_fragment"
     misc_feature    207760. .270196
                     /note="assembly_fragment"
     misc_feature    270297. .270795
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69161 a  65452 c  65038 g  69839 t   1305 others
ORIGIN

  Query Match             43.7%;  Score 30.6;  DB 2;  Length 270795;
  Best Local Similarity   80.0%;  Pred. No. 13;
  Matches   36;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      25 TGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           || | ||| |  |||||| ||||||||||||||||||| ||  ||
Db   29226 TGCATATCCGCGCCCTTCCCGGGAAATGGAAACTGAAAGTCCCAT 29270


RESULT 8
AC022602/c
LOCUS       AC022602               62795 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-700J23 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC022602
VERSION     AC022602.4  GI:13606014
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 62795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-700J23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 12, 2001 this sequence version replaced gi:11990712.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4246
                Center clone name: 700_J_23
            --------------.
            * NOTE: This record contains 73 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      766: contig of 766 bp in length
            *     767 866: gap of      100 bp
            *     867     1645: contig of 779 bp in length
            *    1646 1745: gap of      100 bp
            *    1746     2502: contig of 757 bp in length
            *    2503 2602: gap of      100 bp
            *    2603     3357: contig of 755 bp in length
            *    3358 3457: gap of      100 bp
            *    3458     4231: contig of 774 bp in length
            *    4232 4331: gap of      100 bp
            *    4332     5093: contig of 762 bp in length
            *    5094 5193: gap of      100 bp
            *    5194     5951: contig of 758 bp in length
            *    5952 6051: gap of      100 bp
            *    6052     6877: contig of 826 bp in length
            *    6878 6977: gap of      100 bp
            *    6978     7753: contig of 776 bp in length
            *    7754 7853: gap of      100 bp
            *    7854     8481: contig of 628 bp in length
            *    8482 8581: gap of      100 bp
            *    8582     9379: contig of 798 bp in length
            *    9380 9479: gap of      100 bp
            *    9480    10244: contig of 765 bp in length
            *   10245 10344: gap of      100 bp
            *   10345    11101: contig of 757 bp in length
            *   11102 11201: gap of      100 bp
            *   11202    11966: contig of 765 bp in length
            *   11967 12066: gap of      100 bp
            *   12067    12743: contig of 677 bp in length
            *   12744 12843: gap of      100 bp
```

```
            *    45876    45975: gap of unknown length
            *    45976    47987: contig of 2012 bp in length
            *    47988    48087: gap of unknown length
            *    48088    49414: contig of 1327 bp in length
            *    49415    49514: gap of unknown length
            *    49515    51531: contig of 2017 bp in length
            *    51532    51631: gap of unknown length
            *    51632    53642: contig of 2011 bp in length
            *    53643    53742: gap of unknown length
            *    53743    55270: contig of 1528 bp in length
            *    55271    55370: gap of unknown length
            *    55371    56610: contig of 1240 bp in length
            *    56611    56710: gap of unknown length
            *    56711    58184: contig of 1474 bp in length
            *    58185    58284: gap of unknown length
            *    58285    60285: contig of 2001 bp in length
            *    60286    60385: gap of unknown length
            *    60386    62594: contig of 2209 bp in length
            *    62595    62694: gap of unknown length
            *    62695    64612: contig of 1918 bp in length
            *    64613    64712: gap of unknown length
            *    64713    66753: contig of 2041 bp in length
            *    66754    66853: gap of unknown length
            *    66854    68955: contig of 2102 bp in length
            *    68956    69055: gap of unknown length
            *    69056    71312: contig of 2257 bp in length
            *    71313    71412: gap of unknown length
            *    71413    74128: contig of 2716 bp in length
            *    74129    74228: gap of unknown length
            *    74229    76418: contig of 2190 bp in length
            *    76419    76518: gap of unknown length
            *    76519    78382: contig of 1864 bp in length
            *    78383    78482: gap of unknown length
            *    78483    82106: contig of 3624 bp in length
            *    82107    82206: gap of unknown length
            *    82207    84094: contig of 1888 bp in length
            *    84095    84194: gap of unknown length
            *    84195    85830: contig of 1636 bp in length
            *    85831    85930: gap of unknown length
            *    85931    87937: contig of 2007 bp in length
            *    87938    88037: gap of unknown length
            *    88038    89596: contig of 1559 bp in length
            *    89597    89696: gap of unknown length
            *    89697    92755: contig of 3059 bp in length
            *    92756    92855: gap of unknown length
            *    92856    94689: contig of 1834 bp in length
            *    94690    94789: gap of unknown length
            *    94790    97368: contig of 2579 bp in length

  Query Match             43.7%;  Score 30.6;  DB 2;  Length 196360;
  Best Local Similarity   80.0%;  Pred. No. 12;
   atches   36;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      21 AAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATC 65
           ||  || | |||    ||| ||||||||||||||||||||||||
Db 195228 AATTTGCATATCCCCGCCCCTCTCGGGAAATGGAAACTGAAAATC 195272


RESULT 7
AC015535
LOCUS       AC015535              270795 bp    DNA     linear   HTG 09-MAY-2001
DEFINITION  Mus musculus clone RP23-39M18, WORKING DRAFT SEQUENCE, 14 unordered
            pieces.
ACCESSION   AC015535
VERSION     AC015535.5  GI:11527474
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 270795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-39M18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 270795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 2, 2000 this sequence version replaced gi:7230304.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L909
                Center clone name: 39_M_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 50% of reads
                Sequencing vector: Plasmid; n/a; 50% of reads
                Chemistry: Dye-primer-amersham; 14% of reads
                Chemistry: Dye-terminator Big Dye; 86% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 265948 bases at least Q40
                Consensus quality: 268070 bases at least Q30
                Consensus quality: 268794 bases at least Q20
                Insert size: 247000; agarose-fp
                Insert size: 269495; sum-of-contigs
                Quality coverage: 9.7 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    51876: contig of 51876 bp in length
            *    51877 51976: gap of      100 bp
            *    51977    52976: contig of 1000 bp in length
            *    52977 53076: gap of      100 bp
            *    53077    54220: contig of 1144 bp in length
            *    54221 54320: gap of      100 bp
            *    54321    55540: contig of 1220 bp in length
            *    55541 55640: gap of      100 bp
            *    55641    56917: contig of 1277 bp in length
            *    56918 57017: gap of      100 bp
            *    57018    58094: contig of 1077 bp in length
            *    58095 58194: gap of      100 bp
            *    58195   130154: contig of 71960 bp in length
            *   130155 130254: gap of      100 bp
            *   130255   132870: contig of 2616 bp in length
            *   132871 132970: gap of      100 bp
            *   132971   143777: contig of 10807 bp in length
            *   143778 143877: gap of      100 bp
            *   143878   156806: contig of 12929 bp in length
            *   156807 156906: gap of      100 bp
            *   156907   171040: contig of 14134 bp in length
            *   171041 171140: gap of      100 bp
```

SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 196360)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196360)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 196360)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 11, 2002 this sequence version replaced gi:17973146.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GHFL
Center clone name: CH230-64F11
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 130985 bases at least Q40
Consensus quality: 137495 bases at least Q30
Consensus quality: 142650 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 78 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1005: contig of 1005 bp in length
* 1006 1105: gap of unknown length
* 1106 2120: contig of 1015 bp in length
* 2121 2220: gap of unknown length
* 2221 3271: contig of 1051 bp in length
* 3272 3371: gap of unknown length
* 3372 4942: contig of 1571 bp in length
* 4943 5042: gap of unknown length
* 5043 6695: contig of 1653 bp in length
* 6696 6795: gap of unknown length
* 6796 8482: contig of 1687 bp in length
* 8483 8582: gap of unknown length
* 8583 9611: contig of 1029 bp in length
* 9612 9711: gap of unknown length
* 9712 10802: contig of 1091 bp in length
* 10803 10902: gap of unknown length
* 10903 11944: contig of 1042 bp in length
* 11945 12044: gap of unknown length
* 12045 13423: contig of 1379 bp in length
* 13424 13523: gap of unknown length
* 13524 14596: contig of 1073 bp in length
* 14597 14696: gap of unknown length
* 14697 16601: contig of 1905 bp in length
* 16602 16701: gap of unknown length
* 16702 17907: contig of 1206 bp in length
* 17908 18007: gap of unknown length
* 18008 19613: contig of 1606 bp in length
* 19614 19713: gap of unknown length
* 19714 20826: contig of 1113 bp in length
* 20827 20926: gap of unknown length
* 20927 22558: contig of 1632 bp in length
* 22559 22658: gap of unknown length
* 22659 24413: contig of 1755 bp in length
* 24414 24513: gap of unknown length
* 24514 25973: contig of 1460 bp in length
* 25974 26073: gap of unknown length
* 26074 27357: contig of 1284 bp in length
* 27358 27457: gap of unknown length
* 27458 28672: contig of 1215 bp in length
* 28673 28772: gap of unknown length
* 28773 30074: contig of 1302 bp in length
* 30075 30174: gap of unknown length
* 30175 31717: contig of 1543 bp in length
* 31718 31817: gap of unknown length
* 31818 33178: contig of 1361 bp in length
* 33179 33278: gap of unknown length
* 33279 34465: contig of 1187 bp in length
* 34466 34565: gap of unknown length
* 34566 37040: contig of 2475 bp in length
* 37041 37140: gap of unknown length
* 37141 38197: contig of 1057 bp in length
* 38198 38297: gap of unknown length
* 38298 40231: contig of 1934 bp in length
* 40232 40331: gap of unknown length
* 40332 42055: contig of 1724 bp in length
* 42056 42155: gap of unknown length
* 42156 43274: contig of 1119 bp in length
* 43275 43374: gap of unknown length
* 43375 45875: contig of 2501 bp in length

```
                     /rpt_family="AluY"
     repeat_region   6346. .6655
                     /rpt_family="AluSx"
     repeat_region   complement(6697. .7077)
                     /rpt_family="L2"
     repeat_region   complement(8155. .8456)
                     /rpt_family="AluSg"
     repeat_region   8561. .8653
                     /rpt_family="MIR"
     repeat_region   complement(8687. .8988)
                     /rpt_family="AluSx"
     repeat_region   9084. .9627
                     /rpt_family="MER4A"
     repeat_region   complement(11824. .12115)
                     /rpt_family="AluSx"
     repeat_region   12216. .12454
                     /rpt_family="MIR"
     repeat_region   12457. .12494
                     /rpt_family="(TC)n"
     repeat_region   12494. .12516
                     /rpt_family="(TA)n"
     repeat_region   complement(12891. .13203)
                     /rpt_family="AluSx"
     repeat_region   13685. .13981
                     /rpt_family="AluSx"
     unsure          13964. .13968
                     /note="<30 quality SNGL region"
     repeat_region   complement(14342. .14638)
                     /rpt_family="AluY"
     repeat_region   complement(15438. .15676)
                     /rpt_family="L2"
     repeat_region   complement(15721. .16045)
                     /rpt_family="L2"
     repeat_region   16060. .16221
                     /rpt_family="L1MC5"
     repeat_region   complement(16245. .16567)
                     /rpt_family="L2"
     repeat_region   complement(16779. .17088)
                     /rpt_family="AluJo"
     repeat_region   complement(17394. .17598)
                     /rpt_family="MIR"
     repeat_region   17632. .17928
                     /rpt_family="AluSx"
     repeat_region   17964. .18259
                     /rpt_family="AluSx"
     repeat_region   18714. .18776
                     /rpt_family="L2"
     repeat_region   complement(18831. .19137)
                     /rpt_family="AluSx"
     repeat_region   complement(19145. .19439)
                     /rpt_family="AluSx"
     repeat_region   19483. .19639
                     /rpt_family="AluJb"
     repeat_region   19640. .19933
                     /rpt_family="AluSx"
     repeat_region   19934. .20105
                     /rpt_family="AluJb"
     repeat_region   20107. .20239
                     /rpt_family="AluSx"
     repeat_region   20240. .20546
                     /rpt_family="AluY"

  Query Match             45.1%;  Score 31.6;  DB 9;  Length 174396;
  Best Local Similarity   74.1%;  Pred. No. 5.9;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      14 GAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAG 67
           |||  |||||||||||| ||| ||||||    ||| ||||||  |||||  | ||
Db   46113 GAATAGGAAACTGAAATTCAAATCCCTGACTGGGGAATGGATTCTGAATCTTAG 46060


RESULT 5
AF481734
```

```
LOCUS       AF481734               13943 bp    DNA     linear   ROD 25-JUN-2002
DEFINITION  Mus musculus 2'-5'olygoadenylate synthetase 1b (Flv) gene,
            Flv-C3H.PRI-Flvr allele, complete cds.
ACCESSION   AF481734
VERSION     AF481734.1  GI:21360848
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 13943)
  AUTHORS   Perelygin,A.A., Scherbik,S.V., Zhulin,I.B., Stockman,B.M., Li,Y.
            and Brinton,M.A.
  TITLE     Positional cloning of the murine flavivirus resistance gene
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9322-9327 (2002)
  MEDLINE   22103633
   PUBMED   12080145
REFERENCE   2  (bases 1 to 13943)
  AUTHORS   Perelygin,A.A., Scherbik,S.V. and Brinton,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Biology, Georgia State University, 24
            Peachtree Center Ave., Atlanta, GA 30303, USA
FEATURES             Location/Qualifiers
     source          1. .13943
                     /organism="Mus musculus"
                     /strain="C3H.PRI-Flvr"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /cell_type="SV40-transformed embryofibroblasts"
     gene            1451. .12996
                     /gene="Flv"
                     /allele="C3H.PRI-Flvr"
     mRNA            join(1451. .1693,3240. .3516,6428. .6612,10018. .10250,
                     10901. .11054,12252. .12996)
                     /gene="Flv"
     CDS             join(1514. .1693,3240. .3516,6428. .6612,10018. .10250,
                     10901. .11054,12252. .12353)
                     /gene="Flv"
                     /codon_start=1
                     /product="2'-5'olygoadenylate synthetase 1b"
                     /protein_id="AAM49738.1"
                     /db_xref="GI:21360849"
                     /translation="MEQDLRSIPASKLDKFIENHLPDTSFCADLREVIDALCALLKDR
                     SFRGPVRRMRASKGVKGKGTTLKGRSDADLVVFLNNLTSFEDQLNQQGVLIKEIKKQL
                     CEVQHERRCGVKFEVHSLRSPNSRALSFKLSAPDLLKEVKFDVLPAYDLLDHLNILKK
                     PNQQFYANLISGRTPPGKEGKLSICFMGLRKYFLNCRPTKLKRLIRLVTHWYQLCKEK
                     LGDPLPPQYALELLTVYAWEYGSRVTKFNTAQGFRTVLELVTKYKQLRIYWTVYYDFR
                     HQEVSEYLHQQLKKDRPVILDPADPTRNIAGLNPKDWRRLAGEAATWLQYPCFKYRDG
                     SPVCSWEVPTEVGVPMKYLFCRIFWLLFWSLFHFIFGKTSSG"
     polyA_signal    12974. .12979
                     /gene="Flv"
     polyA_site      12996
                     /gene="Flv"
BASE COUNT     3434 a   3551 c   3318 g   3640 t
ORIGIN

  Query Match             43.7%;  Score 30.6;  DB 10;  Length 13943;
  Best Local Similarity   80.0%;  Pred. No. 11;
  Matches   36;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      25 TGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAGAT 69
           || | ||| |  |||||| ||||||||||||||||||||| ||  ||
Db   1404 TGCATATCCGCGCCCTTCCCGGGAAATGGAAACTGAAGTCCCAT 1448


RESULT 6
AC098508
LOCUS       AC098508               196360 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-64F11, *** SEQUENCING IN PROGRESS
            ***, 78 unordered pieces.
ACCESSION   AC098508
VERSION     AC098508.3  GI:21729584
KEYWORDS    HTG; HTGS_PHASE1.
```

```
Qy      14 GAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAG 67
           ||| |||||||||| ||| ||||||    ||| ||||||  |||||  | ||
Db  38398 GAATAGGAAACTGAAATTCAAATCCCTGACTGGGGAATGGATTCTGAATCTTAG 38451


RESULT 4
AC104564/c
LOCUS       AC104564               174396 bp    DNA     linear   PRI 23-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-68I3, complete sequence.
ACCESSION   AC104564
VERSION     AC104564.10  GI:22450667
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174396)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-68I3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174396)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 174396)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 174396)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 23, 2002 this sequence version replaced gi:22123485.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21550
                Center clone name: 68_I_3
            --------------.
FEATURES             Location/Qualifiers
     source          1. .174396
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-68I3"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   731. .760
                     /rpt_family="(TCC)n"
     repeat_region   1452. .1758
                     /rpt_family="AluSq"
     repeat_region   1759. .1952
                     /rpt_family="MIR"
     repeat_region   complement(1968. .2144)
                     /rpt_family="AluSg/x"
     repeat_region   complement(2320. .2395)
                     /rpt_family="L2"
     repeat_region   2396. .2751
                     /rpt_family="MLT1F"
     repeat_region   complement(2752. .2976)
                     /rpt_family="AluY"
     repeat_region   2977. .3131
                     /rpt_family="MLT1F"
     repeat_region   complement(3132. .3409)
                     /rpt_family="L2"
     repeat_region   3476. .3636
                     /rpt_family="L2"
     repeat_region   complement(3655. .3756)
                     /rpt_family="FLAM_A"
     repeat_region   3761. .3949
                     /rpt_family="L2"
     repeat_region   complement(3950. .4074)
                     /rpt_family="Charlie8"
     repeat_region   4075. .4364
                     /rpt_family="AluJb"
     repeat_region   complement(4365. .4379)
                     /rpt_family="Charlie8"
     repeat_region   complement(4416. .4758)
                     /rpt_family="L2"
     repeat_region   complement(5487. .5614)
                     /rpt_family="L2"
     repeat_region   5966. .6270
```

```
   Insert size: 158298; sum-of-contigs
   Quality coverag.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1115: contig of 1115 bp in length
*     1116 1215: gap of      100 bp
*     1216     2414: contig of 1199 bp in length
*     2415 2514: gap of      100 bp
*     2515     3561: contig of 1047 bp in length
*     3562 3661: gap of      100 bp
*     3662     4929: contig of 1268 bp in length
*     4930 5029: gap of      100 bp
*     5030     6039: contig of 1010 bp in length
*     6040 6139: gap of      100 bp
*     6140     7469: contig of 1330 bp in length
*     7470 7569: gap of      100 bp
*     7570     9736: contig of 2167 bp in length
*     9737 9836: gap of      100 bp
*     9837    12165: contig of 2329 bp in length
*    12166 12265: gap of      100 bp
*    12266    12623: contig of 358 bp in length
*    12624 12723: gap of      100 bp
*    12724    14465: contig of 1742 bp in length
*    14466 14565: gap of      100 bp
*    14566    15963: contig of 1398 bp in length
*    15964 16063: gap of      100 bp
*    16064    18074: contig of 2011 bp in length
*    18075 18174: gap of      100 bp
*    18175    21150: contig of 2976 bp in length
*    21151 21250: gap of      100 bp
*    21251    22852: contig of 1602 bp in length
*    22853 22952: gap of      100 bp
*    22953    25375: contig of 2423 bp in length
*    25376 25475: gap of      100 bp
*    25476    27959: contig of 2484 bp in length
*    27960 28059: gap of      100 bp
*    28060    30399: contig of 2340 bp in length
*    30400 30499: gap of      100 bp
*    30500    34315: contig of 3816 bp in length
*    34316 34415: gap of      100 bp
*    34416    37936: contig of 3521 bp in length
*    37937 38036: gap of      100 bp
*    38037    40122: contig of 2086 bp in length
*    40123 40222: gap of      100 bp
*    40223    43895: contig of 3673 bp in length
*    43896 43995: gap of      100 bp
*    43996    49451: contig of 5456 bp in length
*    49452 49551: gap of      100 bp
*    49552    54107: contig of 4556 bp in length
*    54108 54207: gap of      100 bp
*    54208    60101: contig of 5894 bp in length
*    60102 60201: gap of      100 bp
*    60202    64961: contig of 4760 bp in length
*    64962 65061: gap of      100 bp
*    65062    75980: contig of 10919 bp in length
*    75981 76080: gap of      100 bp
*    76081    89332: contig of 13252 bp in length
*    89333 89432: gap of      100 bp
*    89433   104747: contig of 15315 bp in length
*   104748 104847: gap of      100 bp
*   104848   120555: contig of 15708 bp in length
*   120556 120655: gap of      100 bp
*   120656   161198: contig of 40543 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161198
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-71A1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1115
                     /note="assembly_fragment"
     misc_feature    1216. .2414
                     /note="assembly_fragment"
     misc_feature    2515. .3561
                     /note="assembly_fragment"
     misc_feature    3662. .4929
                     /note="assembly_fragment"
     misc_feature    5030. .6039
                     /note="assembly_fragment"
     misc_feature    6140. .7469
                     /note="assembly_fragment"
     misc_feature    7570. .9736
                     /note="assembly_fragment"
     misc_feature    9837. .12165
                     /note="assembly_fragment"
     misc_feature    12266. .12623
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    12724. .14465
                     /note="assembly_fragment"
     misc_feature    14566. .15963
                     /note="assembly_fragment"
     misc_feature    16064. .18074
                     /note="assembly_fragment"
     misc_feature    18175. .21150
                     /note="assembly_fragment"
     misc_feature    21251. .22852
                     /note="assembly_fragment"
     misc_feature    22953. .25375
                     /note="assembly_fragment"
     misc_feature    25476. .27959
                     /note="assembly_fragment"
     misc_feature    28060. .30399
                     /note="assembly_fragment"
     misc_feature    30500. .34315
                     /note="assembly_fragment"
     misc_feature    34416. .37936
                     /note="assembly_fragment"
     misc_feature    38037. .40122
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    40223. .43895
                     /note="assembly_fragment"
     misc_feature    43996. .49451
                     /note="assembly_fragment"
     misc_feature    49552. .54107
                     /note="assembly_fragment"
     misc_feature    54208. .60101
                     /note="assembly_fragment"
     misc_feature    60202. .64961
                     /note="assembly_fragment"
     misc_feature    65062. .75980
                     /note="assembly_fragment"
     misc_feature    76081. .89332
                     /note="assembly_fragment"
     misc_feature    89433. .104747
                     /note="assembly_fragment"
     misc_feature    104848. .120555
                     /note="assembly_fragment"
     misc_feature    120656. .161198
                     /note="assembly_fragment"
BASE COUNT    40700 a  38352 c  37941 g  41304 t   2901 others
ORIGIN

  Query Match             45.1%;  Score 31.6;  DB 2;  Length 161198;
  Best Local Similarity   74.1%;  Pred. No. 5.9;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;
```

```
*    27254    28110: contig of 857 bp in length
*    28111 28210: gap of      100 bp
*    28211    29057: contig of 847 bp in length
*    29058 29157: gap of      100 bp
*    29158    30015: contig of 858 bp in length
*    30016 30115: gap of      100 bp
*    30116    30959: contig of 844 bp in length
*    30960 31059: gap of      100 bp
*    31060    31892: contig of 833 bp in length
*    31893 31992: gap of      100 bp
*    31993    32825: contig of 833 bp in length
*    32826 32925: gap of      100 bp
*    32926    33786: contig of 861 bp in length
*    33787 33886: gap of      100 bp
*    33887    34730: contig of 844 bp in length
*    34731 34830: gap of      100 bp
*    34831    35686: contig of 856 bp in length
*    35687 35786: gap of      100 bp
*    35787    36642: contig of 856 bp in length
*    36643 36742: gap of      100 bp
*    36743    37588: contig of 846 bp in length
*    37589 37688: gap of      100 bp
*    37689    38583: contig of 895 bp in length
*    38584 38683: gap of      100 bp
*    38684    39538: contig of 855 bp in length
*    39539 39638: gap of      100 bp
*    39639    40465: contig of 827 bp in length
*    40466 40565: gap of      100 bp
*    40566    41398: contig of 833 bp in length
*    41399 41498: gap of      100 bp
*    41499    42350: contig of 852 bp in length
*    42351 42450: gap of      100 bp
*    42451    43323: contig of 873 bp in length
*    43324 43423: gap of      100 bp
*    43424    44274: contig of 851 bp in length
*    44275 44374: gap of      100 bp
*    44375    45243: contig of 869 bp in length
*    45244 45343: gap of      100 bp
*    45344    46198: contig of 855 bp in length
*    46199 46298: gap of      100 bp
*    46299    47210: contig of 912 bp in length
*    47211 47310: gap of      100 bp
*    47311    48149: contig of 839 bp in length
*    48150 48249: gap of      100 bp
*    48250    49095: contig of 846 bp in length
*    49096 49195: gap of      100 bp
*    49196    50025: contig of 830 bp in length
*    50026 50125: gap of      100 bp
*    50126    50955: contig of 830 bp in length
*    50956 51055: gap of      100 bp
*    51056    51889: contig of 834 bp in length
*    51890 51989: gap of      100 bp
*    51990    52860: contig of 871 bp in length
*    52861 52960: gap of      100 bp
*    52961    53793: contig of 833 bp in length
*    53794 53893: gap of      100 bp
*    53894    54738: contig of 845 bp in length
*    54739 54838: gap of      100 bp
*    54839    55682: contig of 844 bp in length
*    55683 55782: gap of      100 bp
*    55783    56640: contig of 858 bp in length
*    56641 56740: gap of      100 bp
*    56741    57576: contig of 836 bp in length
*    57577 57676: gap of      100 bp
*    57677    58527: contig of 851 bp in length
*    58528 58627: gap of      100 bp
*    58628    59468: contig of 841 bp in length
*    59469 59568: gap of      100 bp
*    59569    60400: contig of 832 bp in length
*    60401 60500: gap of      100 bp
*    60501    61390: contig of 890 bp in length
*    61391 61490: gap of      100 bp
*    61491    62346: contig of 856 bp in length
*    62347 62446: gap of      100 bp
*    62447    63310: contig of 864 bp in length
*    63311 63410: gap of      100 bp
*    63411    64262: contig of 852 bp in length
*    64263 64362: gap of      100 bp
*    64363    65192: contig of 830 bp in length
*    65193 65292: gap of      100 bp

  Query Match             45.1%;  Score 31.6;  DB 2;  Length 69861;
  Best Local Similarity   74.1%;  Pred. No. 5.7;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      14 GAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGAAAATCAG 67
           |||  |||||||||||| ||| ||||||    ||| ||||||  |||||  | ||
Db    1327 GAATAGGAAACTGAAATTCAAATCCCTGACTGGGGAATGGATTCTGAATCTTAG 1380


RESULT 3
AC015867
LOCUS       AC015867              161198 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-71A1, WORKING DRAFT SEQUENCE, 30 unordered
            pieces.
ACCESSION   AC015867
VERSION     AC015867.2  GI:7249301
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161198)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-71A1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161198)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6446801.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L641
                Center clone name: 71_A_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                0.2732240437158 47Chemistry: Dye-primer-amersham; 9% of reads
                Chemistry: Dye-terminator Big Dye; 91% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147496 bases at least Q40
                Consensus quality: 154226 bases at least Q30
                Consensus quality: 156653 bases at least Q20
                Insert size: 137000; agarose-fp
```

```
            gene expression with the option of cell growth control
  JOURNAL   Patent: WO 0065074-A 2 02-NOV-2000;
            Gesellschaft f]r Biotechnologische Forschung mbH (GBF) (; DE)
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="Mesocricetus sp."
                     /db_xref="taxon:10038"
                     /note="baby hamster kidney cells"
BASE COUNT       25 a     15 c     15 g     15 t
ORIGIN

  Query Match             100.0%;  Score 70;  DB 6;  Length 70;
  Best Local Similarity   100.0%;  Pred. No. 2e-12;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GATCCCTTCTCGGGAAATGGAAACTGAAAATCAGATCCCTTCTCGGGAAATGGAAACTGA 60

Qy      61 AAATCAGATC 70
           ||||||||||
Db      61 AAATCAGATC 70


RESULT 2
AC032033
LOCUS       AC032033               69861 bp    DNA     linear   HTG 03-APR-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-392F20 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC032033
VERSION     AC032033.1  GI:7387453
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-392F20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu

            -------------- Project Information
                Center project name: L9119
                Center clone name: 392_F_20
            --------------.
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      848: contig of 848 bp in length
            *      849 948: gap of      100 bp
            *      949     1814: contig of 866 bp in length
            *     1815 1914: gap of      100 bp
            *     1915     2781: contig of 867 bp in length
            *     2782 2881: gap of      100 bp
            *     2882     3726: contig of 845 bp in length
            *     3727 3826: gap of      100 bp
            *     3827     4682: contig of 856 bp in length
            *     4683 4782: gap of      100 bp
            *     4783     5612: contig of 830 bp in length
            *     5613 5712: gap of      100 bp
            *     5713     6560: contig of 848 bp in length
            *     6561 6660: gap of      100 bp
            *     6661     7508: contig of 848 bp in length
            *     7509 7608: gap of      100 bp
            *     7609     8438: contig of 830 bp in length
            *     8439 8538: gap of      100 bp
            *    .8539     9375: contig of 837 bp in length
            *     9376 9475: gap of      100 bp
            *     9476    10341: contig of 866 bp in length
            *    10342 10441: gap of      100 bp
            *    10442    11277: contig of 836 bp in length
            *    11278 11377: gap of      100 bp
            *    11378    12218: contig of 841 bp in length
            *    12219 12318: gap of      100 bp
            *    12319    13169: contig of 851 bp in length
            *    13170 13269: gap of      100 bp
            *    13270    13976: contig of 707 bp in length
            *    13977 14076: gap of      100 bp
            *    14077    14931: contig of 855 bp in length
            *    14932 15031: gap of      100 bp
            *    15032    15872: contig of 841 bp in length
            *    15873 15972: gap of      100 bp
            *    15973    16802: contig of 830 bp in length
            *    16803 16902: gap of      100 bp
            *    16903    17723: contig of 821 bp in length
            *    17724 17823: gap of      100 bp
            *    17824    18674: contig of 851 bp in length
            *    18675 18774: gap of      100 bp
            *    18775    19632: contig of 858 bp in length
            *    19633 19732: gap of      100 bp
            *    19733    20580: contig of 848 bp in length
            *    20581 20680: gap of      100 bp
            *    20681    21532: contig of 852 bp in length
            *    21533 21632: gap of      100 bp
            *    21633    22477: contig of 845 bp in length
            *    22478 22577: gap of      100 bp
            *    22578    23424: contig of 847 bp in length
            *    23425 23524: gap of      100 bp
            *    23525    24327: contig of 803 bp in length
            *    24328 24427: gap of      100 bp
            *    24428    25243: contig of 816 bp in length
            *    25244 25343: gap of      100 bp
            *    25344    26186: contig of 843 bp in length
            *    26187 26286: gap of      100 bp
            *    26287    27153: contig of 867 bp in length
            *    27154 27253: gap of      100 bp
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:16 ; Search time 385.676 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: US-09-980-548-2
Perfect score: 70
Sequence: 1 gatcccttctcgggaaatgg.....tggaaactgaaaatcagatc 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		70	100.0	70	6	AX040910	AX040910 Sequence
2		31.6	45.1	69861	2	AC032033	AC032033 Homo sapi
3		31.6	45.1	161198	2	AC015867	AC015867 Homo sapi
4	c	31.6	45.1	174396	9	AC104564	AC104564 Homo sapi
5		30.6	43.7	13943	10	AF481734	AF481734 Mus muscu
6		30.6	43.7	196360	2	AC098508	AC098508 Rattus no
7		30.6	43.7	270795	2	AC015535	AC015535 Mus muscu
8	c	30	42.9	62795	2	AC022602	AC022602 Homo sapi
9	c	30	42.9	148555	2	AP004711	AP004711 Homo sapi
10	c	30	42.9	160883	9	AC021498	AC021498 Homo sapi
11		30	42.9	197743	10	AL591711	AL591711 Mouse DNA
12	c	29.6	42.3	90747	2	AC011975	AC011975 Homo sapi
13	c	29.2	41.7	110121	9	AL590989	AL590989 Human DNA
14	c	29.2	41.7	176775	2	AC016978	AC016978 Homo sapi
15		29	41.4	165373	9	AC067957	AC067957 Homo sapi
16	c	29	41.4	173874	2	AC067791	AC067791 Homo sapi
17		28.8	41.1	61802	2	OSIG00028	AL627350 Oryza sat
18		28.8	41.1	141128	8	OSJN00139	AL662946 Oryza sat
19		28.6	40.9	66317	2	AC110521	AC110521 Mus muscu
20	c	28.6	40.9	213937	10	AL672128	AL672128 Mouse DNA
21	c	28.4	40.6	264461	2	AC130804	AC130804 Medicago
22	c	28.2	40.3	70522	2	AC130491	AC130491 Homo sapi
23	c	28.2	40.3	78603	9	HS1188A21	AL117373 Human DNA
24	c	28.2	40.3	176774	2	AP002519	AP002519 Homo sapi
25	c	28	40.0	31976	9	AL596107	AL596107 Human DNA
26		28	40.0	120724	9	CNS05TCR	AL355837 Human chr
27		28	40.0	144223	9	AC062015	AC062015 Homo sapi
28		28	40.0	145458	2	AC069591	AC069591 Homo sapi
29	c	28	40.0	197659	9	AC009245	AC009245 Homo sapi
30		27.8	39.7	125590	9	AC002382	AC002382 Human BAC
31	c	27.8	39.7	141905	2	AC110101	AC110101 Rattus no
32		27.8	39.7	161831	2	AC073078	AC073078 Homo sapi
33	c	27.8	39.7	163567	2	AC120955	AC120955 Rattus no
34	c	27.8	39.7	180318	2	AC109991	AC109991 Rattus no
35	c	27.8	39.7	186862	2	AC110483	AC110483 Rattus no
36	c	27.8	39.7	214617	2	AC127366	AC127366 Mus muscu
37		27.8	39.7	214934	2	AC068907	AC068907 Mus muscu
38		27.8	39.7	235079	2	AC124475	AC124475 Mus muscu
39	c	27.6	39.4	2352	3	DAHSOP70	X78403 D.auraria m
40	c	27.6	39.4	136641	2	AC114327	AC114327 Canis fam
41		27.6	39.4	168703	4	AC091759	AC091759 Sus scrof
42		27.6	39.4	172884	2	AC113567	AC113567 Canis fam
43		27.6	39.4	184234	2	AC091758	AC091758 Sus scrof
44	c	27.6	39.4	184509	2	AC113237	AC113237 Canis fam
45		27.4	39.1	26483	2	AC100441	AC100441 Mus muscu

ALIGNMENTS

RESULT 1
AX040910
LOCUS AX040910 70 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 2 from Patent WO0065074.
ACCESSION AX040910
VERSION AX040910.1 GI:11340532
KEYWORDS .
SOURCE Mesocricetus sp.
ORGANISM Mesocricetus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Mesocricetus.
REFERENCE 1 (bases 1 to 70)
AUTHORS Mueller,P., Geserick,C., Schroeder,K. and Hauser,H.
TITLE Promoter-transactivator system for inducible high-level mammalian

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11498 row: e column: 22
High quality sequence stop: 318.

FEATURES Location/Qualifiers
source 1. .787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5198973"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."
E COUNT 207 a 274 c 171 g 135 t
ORIGIN

Query Match 21.8%; Score 58.8; DB 13; Length 787;
Best Local Similarity 96.8%; Pred. No. 2.4e-05;
Matches 60; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    209 CCCCGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCCGCCCTCAGCAGTTTCTA 268
          ||| ||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db      7 CCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTA 66

Qy    269 GA 270
          ||
Db     67 GA 68
```

RESULT 15
BB762345
LOCUS BB762345 535 bp mRNA linear EST 17-OCT-2001
DEFINITION BB762345 RIKEN full-length enriched, B16 F10Y cells Mus musculus cDNA clone G370008H11 3', mRNA sequence.
ACCESSION BB762345
VERSION BB762345.1 GI:16196225
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 535)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .535
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370008H11"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"
/note="pooled tissues ; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"
BASE COUNT 172 a 134 c 120 g 109 t
ORIGIN

Query Match 21.7%; Score 58.6; DB 10; Length 535;
Best Local Similarity 65.6%; Pred. No. 2.5e-05;
Matches 120; Conservative 0; Mismatches 54; Indels 9; Gaps 2;

```
Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||| ||  ||||||||||||||||||||||||||||||||   || | ||||   |
Db    111 GCTAACTGCAGTAACGCCATTTTGCAAGGCATGGGAAAATACCAGAGCTGATGTTCTCAG 170

Qy     61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATAT--------GGGCCAAACAGG 112
          |      | |     | | |||||||  | ||| | ||         ||||||||||
Db    171 AAAAACAAGAACAAGGAAGTACAGAGAGGCTGGAAAGTACCGGGACTAGGGCCAAACAGG 230

Qy    113 ATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAA 171
          |||||||||| ||||| |     ||| || ||||||||||||||| |||  |     ||
Db    231 ATATCTGTGGTCAAGCACTAGGGCCCCGGCCCAGGGCCAAGAACAGATGGTCCCCAGAAA 290

Qy    172 TTG 174
          | |
Db    291 TAG 293
```

Search completed: March 6, 2003, 00:22:16
Job time : 2044.42 secs

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 134 a 127 c 98 g 97 t
ORIGIN

Query Match 21.9%; Score 59; DB 10; Length 456;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||  |||    || | |||||     | |
Db      22 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 81

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT------------ATGGGCCAAA 108
           ||||   |  ||| | ||  | | | ||  |||   |             |  |||||||
Db      82 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 141

Qy     109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           ||||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     142 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 198
```

RESULT 13
BB756171
LOCUS BB756171 506 bp mRNA linear EST 17-OCT-2001
DEFINITION BB756171 RIKEN full-length enriched, melanocyte Mus musculus cDNA clone G270072B14 3', mRNA sequence.
ACCESSION BB756171
VERSION BB756171.1 GI:16202641
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 506)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T. , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T. , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .506
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270072B14"
/clone_lib="RIKEN full-length enriched, melanocyte"
/cell_type="melanocyte"
/note="pooled tissues ; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"

BASE COUNT 153 a 130 c 112 g 111 t
ORIGIN

Query Match 21.9%; Score 59; DB 10; Length 506;
Best Local Similarity 67.2%; Pred. No. 2e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||  |||    || | |||||     | |
Db      72 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 131

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT------------ATGGGCCAAA 108
           ||||   |  ||| | ||  | | | ||  |||   |             |  |||||||
Db     132 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 191

Qy     109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           ||||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     192 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 248
```

RESULT 14
BI753609
LOCUS BI753609 787 bp mRNA linear EST 25-SEP-2001
DEFINITION 603028683F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5198973 5', mRNA sequence.
ACCESSION BI753609
VERSION BI753609.1 GI:15745187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

```
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="E860021A18"
                     /clone_lib="RIKEN full-length enriched, 8 cells embryo"
                     /cell_type="8 cells"
                     /dev_stage="8 cells embryo"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      134 a    127 c     98 g     97 t
ORIGIN

  Query Match             21.9%;  Score 59;  DB 10;  Length 456;
  Best Local Similarity   67.2%;  Pred. No. 1.9e-05;
  Matches  119;  Conservative    0;  Mismatches   45;  Indels   13;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||  |||    || | |||||     | |
Db      22 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 81

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT-----------ATGGGCCAAA 108
           ||||   |  ||| | ||  | | | ||  |||   |             |  |||||||
Db      82 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 141

Qy     109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||||  |||||  |    ||| ||||||||||||||||||| ||| ||
Db     142 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 198


RESULT 12
BB733115
LOCUS       BB733115                456 bp    mRNA    linear   EST 12-OCT-2001
DEFINITION  BB733115 RIKEN full-length enriched, 12 days embryo whole body Mus
            musculus cDNA clone E970035B21 3', mRNA sequence.
ACCESSION   BB733115
VERSION     BB733115.1  GI:16116390
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="E970035B21"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     whole body"
                     /tissue_type="whole body"
                     /dev_stage="12 days embryo"
```

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 131 a 122 c 96 g 96 t
ORIGIN

Query Match 21.9%; Score 59; DB 10; Length 445;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||||  |||    || | |||||     | |
Db      12 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 71

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT------------ATGGGCCAAA 108
           ||||   |  ||| | ||  | | | ||  |||   |             | |||||||
Db      72 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 131

Qy     109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     132 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 188
```

RESULT 10
BB670350
LOCUS BB670350 455 bp mRNA linear EST 05-OCT-2001
DEFINITION BB670350 RIKEN full-length enriched, 17 days embryo head Mus musculus cDNA clone 3322402I23 3', mRNA sequence.
ACCESSION BB670350
VERSION BB670350.1 GI:15969571
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 455)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .455
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3322402I23"
/clone_lib="RIKEN full-length enriched, 17 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="17 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 134 a 126 c 98 g 97 t
ORIGIN

Query Match 21.9%; Score 59; DB 10; Length 455;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||||  |||    || | |||||     | |
Db      21 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 80

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT------------ATGGGCCAAA 108
           ||||   |  ||| | ||  | | | ||  |||   |             | |||||||
Db      81 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 140

Qy     109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     141 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 197
```

RESULT 11
BB726927
LOCUS BB726927 456 bp mRNA linear EST 12-OCT-2001
DEFINITION BB726927 RIKEN full-length enriched, 8 cells embryo Mus musculus cDNA clone E860021A18 3', mRNA sequence.
ACCESSION BB726927
VERSION BB726927.1 GI:16110202
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 456)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,

Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

'URES Location/Qualifiers
source 1. .439
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="E970034P16"
/clone_lib="RIKEN full-length enriched, 12 days embryo
whole body"
/tissue_type="whole body"
/dev_stage="12 days embryo"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearrayed set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5 embryos
, and E12.5 female mesonephros/gonad) and one newborn
ovary cDNA library. Average insert size 1.5 kb. All
source libraries are cloned unidirectionally with Oligo(dT
)-NotI primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray, 2000,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse develolpment, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
Mol Genet 7: 1967-1978."
; COUNT 130 a 120 c 93 g 96 t
ORIGIN

Query Match 21.9%; Score 59; DB 10; Length 439;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

```
Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||  |||||||||||||||||||||||||  |||    || | |||||     | |
Db      6 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 65

Qy     61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAAT------------ATGGGCCAAA 108
          ||||   |  ||| | ||  | | | ||  |||   |             |  |||||||
Db     66 AGTTACAAGGAAGTTTAGTTAAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAA 125

Qy    109 CAGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
          |||||||||||||| |||||  |    ||| ||||||||||||||||| ||| ||
Db    126 CAGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 182
```

RESULT 9

BB670654
LOCUS BB670654 445 bp mRNA linear EST 05-OCT-2001
DEFINITION BB670654 RIKEN full-length enriched, 17 days embryo head Mus musculus cDNA clone 3322404N14 3', mRNA sequence.
ACCESSION BB670654
VERSION BB670654.1 GI:15969875
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 445)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T. , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T. , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .445
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3322404N14"
/clone_lib="RIKEN full-length enriched, 17 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="17 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G930034B02"
/clone_lib="RIKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA"
/tissue_type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
/note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1), (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line= CRL-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=CCL-142 RAG), (tissue_type=submandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)"
BASE COUNT 133 a 130 c 98 g 85 t
ORIGIN

Query Match 22.8%; Score 61.6; DB 10; Length 446;
Best Local Similarity 63.2%; Pred. No. 3.9e-06;
Matches 115; Conservative 0; Mismatches 59; Indels 8; Gaps 1;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||| ||  |||||||||||| ||||||||||||||||||   || | ||||   |
Db      25 GCTAACTCCAGTAACGCCATCTTGCAAGGCATGGGAAAATACCAGAGCTGATGTTCTCAG 84

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATAT--------GGGCCAAACAGG 112
           |      | |     | | |||||||  | ||| | ||         ||||||||||||
Db      85 AAAAACAAGAACAAGGAAGTACAGAGAGGCTGGAAAGTACCGGGACTAGGGCCAAACAGG 144

Qy     113 ATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT 172
           |||||||||||| | |   |   ||||||  |  ||||||||||| |||  |     |||
Db     145 ATATCTGTGGTCAACCACTAGGGCCCCGGCCCAGGCCAAGAACAGATGGTCCCCAGAAAT 204

Qy     173 TG 174
           ||
Db     205 AG 206
```

RESULT 7
AW750456/c
LOCUS AW750456 284 bp mRNA linear EST 28-APR-2000
DEFINITION RC5-BT0604-260100-011-C11 BT0604 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW750456
VERSION AW750456.1 GI:7665388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 284)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC5&t2=RC5-BT0604-260100-011-C11&t3=2000-01-26&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 283.
FEATURES Location/Qualifiers
source 1. .284
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0604"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 45 a 76 c 65 g 98 t
ORIGIN

Query Match 22.0%; Score 59.4; DB 10; Length 284;
Best Local Similarity 67.4%; Pred. No. 1.4e-05;
Matches 118; Conservative 0; Mismatches 46; Indels 11; Gaps 2;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||  ||||||||||||||||||||||||  |||    || | |||||     | |
Db     220 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 161

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAAC----------AGGAGAATATGGGCCAAACA 110
           |||     |  ||| | ||  | |||  ||           |   | | | ||||||||||
Db     160 AGTCACAAGGAAGTTTAGTTAAAGAATAAGGCTGAACAAAACTGGGACAGGGGCCAAACA 101

Qy     111 GGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||  ||||  |    ||| ||||||||||||||||| ||| ||
Db     100 GGATATCTGTGGTCGAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 46
```

RESULT 8
BB733109
LOCUS BB733109 439 bp mRNA linear EST 12-OCT-2001
DEFINITION BB733109 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970034P16 3', mRNA sequence.
ACCESSION BB733109
VERSION BB733109.1 GI:16116384
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001).
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

```
RESULT 5
BB750356
LOCUS       BB750356                444 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  BB750356 RIKEN full-length enriched, pooled tissues, cerebellum,
            etc. Mus musculus cDNA clone G130003D06 3', mRNA sequence.
ACCESSION   BB750356
VERSION     BB750356.1  GI:16154592
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G130003D06"
                     /clone_lib="RIKEN full-length enriched, pooled tissues,
                     cerebellum, etc."
                     /note="pooled tissues ; (tissue_type=cerebellum,
                     dev_stage=16 days neonate, sex=mixed),
                     (tissue_type=cerebellum, dev_stage=0 day neonate,
                     sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
                     sex=male), (tissue_type=whole body, dev_stage=9 days
                     embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
                     embryo, sex=mixed)"
BASE COUNT      129 a    120 c     99 g     96 t
ORIGIN
```

```
  Query Match           22.8%;  Score 61.6;  DB 10;  Length 444;
  Best Local Similarity  68.2%;  Pred. No. 3.9e-06;
  Matches  120;  Conservative    0;  Mismatches   44;  Indels   12;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           | |||||  |||||||||||||||||||||||  ||||   || | |||||     | |
Db      11 GGTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAATACCAGAGCTGAGTTCTCAAA 70

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATG----------GGCCAAAC 109
           ||||   |  ||| | ||  | | | | |  |  |||||           ||||||||
Db      71 AGTTACAAGGAAGTTTAGTTAAAAAAATAAGGCTGAATAATACTGAGACAGAGGCCAAAC 130

Qy     110 AGGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     131 AGGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 186
```

```
RESULT 6
BB835008
LOCUS       BB835008                446 bp    mRNA    linear   EST 19-NOV-2001
DEFINITION  BB835008 RIKEN full-length enriched, mammary gland RCB-0527
            Jyg-MC(B) cDNA Mus musculus cDNA clone G930034B02 3', mRNA
            sequence.
ACCESSION   BB835008
VERSION     BB835008.1  GI:17013251
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .446
```

```
LOCUS       BF101077                1240 bp    mRNA    linear   EST 19-OCT-2000
DEFINITION  601754641F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3983531 5',
            mRNA sequence.
ACCESSION   BF101077
VERSION     BF101077.1  GI:10883603
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1240)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9184  row: b  column: 12
            High quality sequence stop: 580.
FEATURES             Location/Qualifiers
     source          1. .1240
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3983531"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      472 a    310 c    308 g    149 t      1 others
ORIGIN

  Query Match             23.2%;  Score 62.6;  DB 12;  Length 1240;
  Best Local Similarity   68.6%;  Pred. No. 2.5e-06;
  Matches  120;  Conservative    0;  Mismatches   44;  Indels   11;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||||  |||    || | |||||     | |
Db     148 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 207

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAAC----------AGGAGAATATGGGCCAAACA 110
           ||||   |  ||| | ||  | |||  ||           |   | | | ||||||||||
Db     208 AGTTACAAGGAAGTTTAGTTAAAGATTAAGGCTGAATAATACCGGGACAGGGGCCAAACA 267

Qy     111 GGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           ||||||||||||| |||||  |    ||| ||||||||||||||||||| ||| ||
Db     268 GGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 322


RESULT 4
AZ804164/c
LOCUS       AZ804164                666 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0064J24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0064J24 R, DNA sequence.
ACCESSION   AZ804164
VERSION     AZ804164.1  GI:12956487
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0064  row: J  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0064J24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      141 a    166 c    148 g    211 t
ORIGIN

  Query Match             22.9%;  Score 61.8;  DB 17;  Length 666;
  Best Local Similarity   58.1%;  Pred. No. 3.7e-06;
  Matches  155;  Conservative    0;  Mismatches   97;  Indels   15;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  ||||||||||||||||||||||||  |||    || | |||||     | |
Db     335 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 276

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           ||||   |  ||| ||||  | |||  |||||   || ||   | |  | | |||    |
Db     275 AGTTACAAGGAAGTTCAGTTAAAGATTAACAGTTAAAAAT---CAAGGCTGAATAATACT 219

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
           | |   |  |  |  ||| |   | |||           | ||     |  | |
Db     218 AGGACAAGGGCCAAGAACCGATGGTGCCCAC-----------CTGGGCCCAGGATCAAG 171

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
           |     | | ||||||  ||  |    ||| ||||||||||||||||||||||| ||||
Db     170 GCCAAGAACCGGTGGTACCCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTACCCA 111

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCT 267
           |||  ||     ||  || ||||||||
Db     110 GATAAGGCGGAACCAGCAACAGTTTCT 84
```

Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

```
FURES             Location/Qualifiers
  source          1. .538
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="G270003K01"
                  /clone_lib="RIKEN full-length enriched, melanocyte"
                  /cell_type="melanocyte"
                  /note="pooled tissues ; (tissue_type=cerebellum,
                  dev_stage=16 days neonate, sex=mixed),
                  (tissue_type=cerebellum, dev_stage=0 day neonate,
                  sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
                  sex=male), (tissue_type=whole body, dev_stage=9 days
                  embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
                  embryo, sex=mixed)"
BASE COUNT      172 a    137 c    117 g    112 t
ORIGIN

  Query Match             23.4%;  Score 63.2;  DB 10;  Length 538;
  Best Local Similarity   63.7%;  Pred. No. 1.5e-06;
  Matches  116;  Conservative    0;  Mismatches   58;  Indels    8;  Gaps    1;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||| ||  |||||||||| |||||||||||||||||   || | || |   |
Db     116 GCTAACTGCAGTAACGCCATCTTGCAAGGCATGGGAAAATACCAGAGCTTATTTTCTCAG 175

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATAT--------GGGCCAAACAGG 112
           |     |  |     | | |||||||  | ||| | ||        ||||||||||||
       176 AAAAACAAGAACAAGGAAGTACAGAGAGGCTGGAAAGTACCGGGAGTAGGGCCAAACAGG 235

Qy     113 ATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT 172
           |||||||||||  |     |   ||||||| |||||||||||||| |||  |     |||
Db     236 ATATCTGTGGTCAAGCACTAGGGCCCCGCCCAGGGCCAAGAACAGATGGTCCCCAGAAAT 295

Qy     173 TG 174
           ||
Db     296 AG 297


RESULT 2
AZ984553
LOCUS       AZ984553                616 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0266C09F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0266C09 F, DNA sequence.
ACCESSION   AZ984553
VERSION     AZ984553.1  GI:13855780
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0266  row: C  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 616.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0266C09"
                      /clone_lib="Mouse 10kb plasmid UUGC2M library"
                      /sex="Female"
                      /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (female) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      180 a    151 c    141 g    144 t
ORIGIN

  Query Match             23.2%;  Score 62.6;  DB 17;  Length 616;
  Best Local Similarity   68.6%;  Pred. No. 2.2e-06;
  Matches  120;  Conservative    0;  Mismatches   44;  Indels   11;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||  |||||||||||||||||||||||||  |||    || | |||||     | |
Db     225 GCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA 284

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAAC----------AGGAGAATATGGGCCAAACA 110
           ||||   |  ||| | ||  | |||  ||           |   | | | ||||||||||
Db     285 AGTTACAAGGAAGTTTAGTTAAAGAATAAGGCTGAATAATACCGGGACAGGGGCCAAACA 344

Qy     111 GGATATCTGTGGT-AAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAAC 164
           |||||||||||||  |||||  |    |||  |||||||||||||||||||  ||| ||
Db     345 GGATATCTGTGGTCAAGCACCTGGGCCCCGGCTCAGGGCCAAGAACAGATGGTAC 399


RESULT 3
BF101077
```

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 2036.09 Seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-1
Perfect score: 270
Sequence: 1 gctagcttaagtaacgccat.....cgccctcagcagtttctaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	63.2	23.4	538	10	BB751692	BB751692 BB751692
	2	62.6	23.2	616	17	AZ984553	AZ984553 2M0266C09
	3	62.6	23.2	1240	12	BF101077	BF101077 601754641
c	4	61.8	22.9	666	17	AZ804164	AZ804164 2M0064J24
	5	61.6	22.8	444	10	BB750356	BB750356 BB750356
	6	61.6	22.8	446	10	BB835008	BB835008 BB835008
c	7	59.4	22.0	284	10	AW750456	AW750456 RC5-BT060
	8	59	21.9	439	10	BB733109	BB733109 BB733109
	9	59	21.9	445	10	BB670654	BB670654 BB670654
	10	59	21.9	455	10	BB670350	BB670350 BB670350
	11	59	21.9	456	10	BB726927	BB726927 BB726927
	12	59	21.9	456	10	BB733115	BB733115 BB733115
	13	59	21.9	506	10	BB756171	BB756171 BB756171
	14	58.8	21.8	787	13	BI753609	BI753609 603028683
	15	58.6	21.7	535	10	BB762345	BB762345 BB762345
	16	58.6	21.7	537	10	BB762595	BB762595 BB762595
c	17	58.6	21.7	584	13	BI794183	BI794183 ic89e10.x
c	18	58.6	21.7	596	13	BI677171	BI677171 ic82f08.x
	19	58.6	21.7	644	10	BB795130	BB795130 BB795130
	20	58.6	21.7	649	10	BB795436	BB795436 BB795436
	21	58.6	21.7	650	10	BB794939	BB794939 BB794939
	22	58.6	21.7	660	10	BB794906	BB794906 BB794906
	23	58.6	21.7	662	10	BB795586	BB795586 BB795586
	24	58.6	21.7	663	10	BB794704	BB794704 BB794704
	25	58.6	21.7	666	10	BB795529	BB795529 BB795529
	26	58.6	21.7	680	10	BB794600	BB794600 BB794600
	27	58.6	21.7	680	10	BB794643	BB794643 BB794643
	28	58.6	21.7	687	10	BB794760	BB794760 BB794760
	29	58.6	21.7	688	10	BB794665	BB794665 BB794665
	30	58.6	21.7	691	10	BB794587	BB794587 BB794587
	31	57.6	21.3	209	9	AA666591	AA666591 vm49c04.r
	32	57.6	21.3	209	9	AA146207	AA146207 mq88d06.r
	33	57.6	21.3	210	9	AA655589	AA655589 vs44h02.r
	34	57.6	21.3	210	9	AA655895	AA655895 vs43g08.r
	35	57.6	21.3	210	9	AA684134	AA684134 vm53b10.r
	36	57.6	21.3	210	9	AA122917	AA122917 mq69d03.r
	37	57.6	21.3	269	9	AA139125	AA139125 mq85c01.r
	38	57.6	21.3	444	10	BB831589	BB831589 BB831589
	39	57.6	21.3	726	12	BF162178	BF162178 601770652
	40	57	21.1	426	10	BB786738	BB786738 BB786738
	41	57	21.1	435	10	BB821235	BB821235 BB821235
	42	57	21.1	444	10	BB788783	BB788783 BB788783
	43	57	21.1	449	10	BB815029	BB815029 BB815029
	44	57	21.1	461	10	BB833539	BB833539 BB833539
	45	57	21.1	525	10	BB762415	BB762415 BB762415

ALIGNMENTS

RESULT 1
BB751692
LOCUS BB751692 538 bp mRNA linear EST 16-OCT-2001
DEFINITION BB751692 RIKEN full-length enriched, melanocyte Mus musculus cDNA clone G270003K01 3', mRNA sequence.
ACCESSION BB751692
VERSION BB751692.1 GI:16155928
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

CC proteins in mammalian tissue culture host cells, including rat fibroblast
CC cells, bovine kidney cells and human kidney cells. The present sequence
CC is cytomegalovirus (CMV) MN14 vector used in the invention. The vector
CC comprises the following elements: CMV promoter, MN14 antibody heavy
CC chain signal peptide, MN14 antibody heavy chain, encephalomyocarditis
CC virus (ECMV) IRES -bovine alpha-lactalbumin signal peptide, MN14 antibody
CC light chain and 3' moloney murine leukemia virus (MoMuLV) LTR.
XX
SQ Sequence 4207 BP; 1030 A; 1202 C; 1058 G; 917 T; 0 other;

Query Match 75.1%; Score 202.8; DB 24; Length 4207;
Best Local Similarity 95.6%; Pred. No. 1.1e-55;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    3644 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 3702

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          ||||||||||||||||||||||||| | |||||  ||||||||||||||||||||||||||
Db    3703 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 3762

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| ||||||||  |||| |||||
Db    3763 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 3822

Qy     178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    3823 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 3882

Qy     237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||| ||||||||||||||||||||
Db    3883 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 3916
```

Search completed: March 5, 2003, 22:49:51
Job time : 277.478 secs

```
CC   two integrated integrating vectors. The integrating vectors comprise at
CC   least one exogenous gene operably linked to a promoter. The host cell
CC   is useful for producing a desired protein and for comparing protein
CC   functions. The host cells comprises a reporter gene which is from gene
CC   fluorescent protein, luciferase, beta-galactosidase and beta-lactamase,
CC   and the assaying step further comprises detecting a signal from the
CC   reporter gene. The desired protein includes proteins for pharmaceutical,
CC   industrial, diagnostic and other purposes. The host cells is useful for
CC   indirectly detecting the expression of a desired protein, comprising
CC   providing the host cell transfected with a vector encoding a
CC   polycistronic sequence comprising a signal protein and a desired protein
CC   operably linked by an internal ribosome entry site (IRES), and culturing
CC   the host cell under suitable conditions so that the signal protein and
CC   the desired protein is produced, where the presence of signal protein
CC   indicates the presence of desired protein. The present sequence is
CC   cytomegalovirus (CMV) MN14 vector used in the invention. The vector
CC   comprises the following elements: CMV promoter, MN14 antibody heavy
CC   chain signal peptide, MN14 antibody heavy chain, encephalomyocarditis
CC   virus (ECMV) IRES -bovine alpha-lactalbumin signal peptide, MN14 antibody
CC   light chain and 3' moloney murine leukemia virus (MoMuLV) LTR.
XX
SQ   Sequence 4207 BP; 1030 A; 1202 C; 1058 G; 917 T; 0 other;

  ery Match             75.1%;  Score 202.8;  DB 24;  Length 4207;
 est Local Similarity   95.6%;  Pred. No. 1.1e-55;
 Matches  262;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    3644 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 3702

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
           |||||||||||||||||||||||| | |||||  ||||||||||||||||||||||||||
Db    3703 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 3762

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
           |||||||||||||||||||| |||||||||||||||||| ||||||||  |||| |||||
Db    3763 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 3822

Qy     178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    3823 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 3882

Qy     237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||||||| ||||||||||||||||||
Db    3883 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 3916


RESULT 15
AAD28306
ID   AAD28306 standard; DNA; 4207 BP.

     AAD28306;
XX
DT   22-APR-2002  (first entry)
XX
DE   Cytomegalovirus (CMV) MN14 vector.
XX
KW   Regulatory element; vector; erythropoietin; growth hormone; insulin;
KW   immunoglobulin; bone morphogenetic protein; interferon; interleukin;
KW   superoxide dismutase; T-cell receptor; surface membrane protein; CMV;
KW   viral antigen; transport protein; addressin; regulatory protein; ECMV;
KW   encephalomyocarditis virus; internal ribosome entry site; IRES; bovine;
KW   cytomegalovirus; moloney murine leukemia virus; MoMuLV; chimeric;
KW   alpha-lactalbumin; promoter; ds.
XX
OS   Chimeric - Encephalomyocarditis virus.
OS   Chimeric - Bos sp.
OS   Chimeric - Cytomegalovirus.
OS   Chimeric - Moloney murine leukemia virus.
OS   Chimeric - Unidentified.
XX
FH   Key             Location/Qualifiers

FT   promoter        1..812
FT                   /*tag= a
FT                   /note= "CMV promoter/enhancer"
FT   misc_signal     853..855
FT                   /*tag= b
FT                   /note= "MN14 antibody heavy chain gene signal peptide
FT                   start codon"
FT   misc_signal     2257..2259
FT                   /*tag= c
FT                   /note= "MN14 antibody heavy chain gene start codon"
FT   misc_feature    2271..2846
FT                   /*tag= d
FT                   /note= "EMCV IRES"
FT   misc_signal     2847..2849
FT                   /*tag= e
FT                   /note= "Bovine alpha-lactalbumin signal peptide start
FT                   codon"
FT   misc_signal     2904..2906
FT                   /*tag= f
FT                   /note= "First codon mature MN14 antibody light chain
FT                   gene"
FT   misc_signal     3543..3544
FT                   /*tag= g
FT                   /note= "MN14 antibody heavy chain gene stop codon"
FT   LTR             3614..4207
FT                   /*tag= h
FT                   /note= "MoMuLV 3' LTR"
XX
PN   WO200202783-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-US20714.
XX
PR   03-JUL-2000; 2000US-215851P.
XX
PA   (GALA-) GALA DESIGN INC.
XX
PI   Bleck GT;
XX
DR   WPI; 2002-154749/20.
XX
PT   Novel regulatory elements including nucleic acid encoding hybrid
PT   alpha-lactalbumin promoter or mutant RNA export element, for expressing
PT   one or more proteins e.g. antibodies, pharmaceutical proteins in host
PT   cells -
XX
PS   Example 1; Fig 7; 151pp; English.
XX
CC   The invention relates to novel regulatory elements and vectors for the
CC   expression of one or more proteins in a host cell. The invention further
CC   provides methods of indirectly detecting the expression of a protein of
CC   interest, comprising providing the host cell transfected with a vector
CC   encoding a polycistronic sequence comprising a signal protein and a
CC   desired protein operably linked by an internal ribosome entry site
CC   (IRES), and culturing the host cell under suitable conditions so that
CC   the signal protein and the desired protein is produced, where the
CC   presence of signal protein indicates the presence of desired protein.
CC   Regulatory elements and vectors of the invention are useful for the
CC   expression of proteins of interest in a host cell. They are useful for
CC   producing an immunoglobulin (Ig), preferably secretory Ig. They are
CC   useful in the expression of one or more proteins such as erythropoietin,
CC   growth hormone, insulin, immunoglobulins, protein C, cytokines and their
CC   receptors, hormones, Von Willebrands factor, lung surfactant, serum
CC   albumins, DNase, vascular endothelial growth factor, receptors for
CC   hormones or growth factors, rheumatoid factors, nerve growth factors,
CC   CD proteins, osteoinductive factors, immunotoxins, bone morphogenetic
CC   protein, interferons, colony stimulating factors, interleukins,
CC   superoxide dismutase, T-cell receptors, surface membrane proteins,
CC   viral antigens, transport proteins, addressins, regulatory proteins,
CC   antibodies, chimeric proteins and their fragments. The vectors are
CC   particularly useful for expressing G protein coupled receptors and other
CC   transmembrane proteins. The retroviral vectors are useful for expressing
```

```
FT                  /*tag= i
FT                  /note= "MoMuLV 3' LTR"
XX
PN   WO200202738-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-US20710.
XX
PR   03-JUL-2000; 2000US-215925P.
XX
PA   (GALA-) GALA DESIGN INC.
XX
PI   Bremel RD, Miller LU, Bleck GT, York D;
XX
DR   WPI; 2002-154737/20.
XX
PT   Host cell for producing a desired protein and for screening compounds
PT   useful for pharmaceutical, industrial, diagnostic and other purposes,
PT   comprises multiple integrating vectors having an exogenous gene -
XX
PS   Example 1; Fig 11; 191pp; English.
XX
CC   The invention relates to a host cell comprising a genome having at least
CC   two integrated integrating vectors. The integrating vectors comprise at
CC   least one exogenous gene operably linked to a promoter. The host cell
CC   is useful for producing a desired protein and for comparing protein
CC   functions. The host cells comprises a reporter gene which is from gene
CC   fluorescent protein, luciferase, beta-galactosidase and beta-lactamase,
CC   and the assaying step further comprises detecting a signal from the
CC   reporter gene. The desired protein includes proteins for pharmaceutical,
CC   industrial, diagnostic and other purposes. The host cells is useful for
CC   indirectly detecting the expression of a desired protein, comprising
CC   providing the host cell transfected with a vector encoding a
CC   polycistronic sequence comprising a signal protein and a desired protein
CC   operably linked by an internal ribosome entry site (IRES), and culturing
CC   the host cell under suitable conditions so that the signal protein and
CC   the desired protein is produced, where the presence of signal protein
CC   indicates the presence of desired protein. The present sequence is
CC   alpha lactalbumin (LA) Bot vector used in the invention. The vector
CC   comprises the following elements: bovine/human alpha-lactalbumin hybrid
CC   promoter, double mutated pre-mRNA processing enhancer (PPE) sequence;
CC   cc49 signal peptide, botulinum toxin antibody light chain, IRES from
CC   encephalomyocarditis virus (ECMV), bovine alpha-lactalbumin signal
CC   peptide, botulinum toxin antibody heavy chain, woodchuck mRNA processing
CC   enhancer (WPRE) sequence and 3' moloney murine leukemia virus (MoMuLV)
CC   LTR.
XX
SQ   Sequence 3671 BP; 907 A; 868 C; 925 G; 971 T; 0 other;

  Query Match             75.1%;  Score 202.8;  DB 24;  Length 3671;
  Best Local Similarity   95.6%;  Pred. No. 1e-55;
  Matches  262;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    3108 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 3166

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||||| | |||||  ||||||||||||||||||||||||
Db    3167 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 3226

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||| ||||||||  |||| |||||
Db    3227 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 3286

Qy     178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    3287 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 3346

Qy     237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||| ||||||||||||||||||||
Db    3347 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 3380

RESULT 14
AAD28267
ID   AAD28267 standard; DNA; 4207 BP.
XX
AC   AAD28267;
XX
DT   22-APR-2002  (first entry)
XX
DE   Cytomegalovirus (CMV) MN14 vector.
XX
KW   Bovine; alpha-lactalbumin; promoter; pharmaceutical; industrial; ECMV;
KW   encephalomyocarditis virus; diagnostic; internal ribosome entry site;
KW   IRES; screening; CMV; cytomegalovirus; moloney murine leukemia virus;
KW   MoMuLV; chimeric; ds.
XX
OS   Chimeric - Encephalomyocarditis virus.
OS   Chimeric - Bos sp.
OS   Chimeric - Cytomegalovirus.
OS   Chimeric - Moloney murine leukemia virus.
OS   Chimeric - Unidentified.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..812
FT                   /*tag= a
FT                   /note= "CMV promoter/enhancer"
FT   misc_signal     853..855
FT                   /*tag= b
FT                   /note= "MN14 antibody heavy chain gene signal peptide
FT                   start codon"
FT   misc_signal     2257..2259
FT                   /*tag= c
FT                   /note= "MN14 antibody heavy chain gene start codon"
FT   misc_feature    2271..2846
FT                   /*tag= d
FT                   /note= "EMCV IRES"
FT   misc_signal     2847..2849
FT                   /*tag= e
FT                   /note= "Bovine alpha-lactalbumin signal peptide start
FT                   codon"
FT   misc_signal     2904..2906
FT                   /*tag= f
FT                   /note= "First codon mature MN14 antibody light chain
FT                   gene"
FT   misc_signal     3543..3544
FT                   /*tag= g
FT                   /note= "MN14 antibody light chain gene stop codon"
FT   LTR             3614..4207
FT                   /*tag= h
FT                   /note= "MoMuLV 3' LTR"
XX
PN   WO200202738-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-US20710.
XX
PR   03-JUL-2000; 2000US-215925P.
XX
PA   (GALA-) GALA DESIGN INC.
XX
PI   Bremel RD, Miller LU, Bleck GT, York D;
XX
DR   WPI; 2002-154737/20.
XX
PT   Host cell for producing a desired protein and for screening compounds
PT   useful for pharmaceutical, industrial, diagnostic and other purposes,
PT   comprises multiple integrating vectors having an exogenous gene -
XX
PS   Example 1; Fig 7; 191pp; English.
XX
CC   The invention relates to a host cell comprising a genome having at least
```

```
Qy    237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||||| ||||||||||||||||||
Db    270 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 303
```

RESULT 12
AAF83096
ID AAF83096 standard; DNA; 3097 BP.
XX
AC AAF83096;
XX
DT 29-JUN-2001 (first entry)
XX
DE Nucleotide sequence of a LTR plasmid.
XX
KW Retrovirus; recombinase recognition sequence; RRS; LTR; recombinase;
KW long terminal repeat; pharmaceutical; cytostatic; antiinflammatory;
KW antirheumatic; antiarthritic; antiasthmatic; osteopathic; cardiant; MLV;
KW vasotropic; neuroprotective; nootropic; cerebroprotective; antipsoriatic;
KW antiarteriosclerotic; vulnerary; anti-HIV; antiulcer; thrombolytic;
KW dermatological; gene therapy; ss.
XX
OS Synthetic.

WO200125466-A1.
XX
PD 12-APR-2001.
XX
PF 05-OCT-2000; 2000WO-GB03837.
XX
PR 05-OCT-1999; 99GB-0023558.
XX
PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI Slingsby J, Kingsman SM, Rohll J, Slade A;
XX
DR WPI; 2001-281732/29.
XX
PT Modifying producer cells for making retrovirus by transfecting with a
PT construct comprising a 5'- recombinase recognition sequence, long
PT terminal repeat and 3'- recombinase recognition sequence, in presence
PT of recombinase -
XX
PS Example 5; Page 129-130; 133pp; English.
XX
CC The invention relates to a method of modifying producer cells for making
CC retrovirus by transfecting with a construct comprising a 5'- recombinase
CC recognition sequence (RRS), long terminal repeat (LTR0 and 3'- RRS, in
CC presence of recombinase. The regulated retroviral vector produced is
CC useful in the manufacture of a pharmaceutical composition to deliver a
CC NOI to a target site, and in the manufacture of a medicament for
CC diagnostic, therapeutic and/or medical applications. The recombinase
CC assisted method is useful for introducing regulated 3'-LTR into a derived
CC producer cell line to produce a high titer regulated retroviral vector.
CC The vector is useful in gene therapy for treating diseases like cancers,
CC inflammatory diseases, immunological disorders such as graft vs host
CC disease, autoimmune diseases such as rheumatoid arthritis, allergic
CC diseases such as asthma, osteoporosis, cardiovascular diseases such as
CC congestive heart failure and ischemic heart disease, neurodegenerative
CC disorders such as multiple sclerosis, Alzheimer's disease, stroke and
CC cerebral ischemia, atherosclerosis, thrombotic disorders, dermatological
CC disorders such as atopic dermatitis, contact dermatitis and psoriasis,
CC wound healing, restenosis, infectious disorders such as HIV infections,
CC ulcers, digestive disorders such as anorexia, bulimia and cachexia, and
CC other diseases. The present sequence represents the nucleotide sequence
CC of a LTR plasmid.
XX
SQ Sequence 3097 BP; 770 A; 823 C; 763 G; 741 T; 0 other;

Query Match 75.1%; Score 202.8; DB 22; Length 3097;
Best Local Similarity 95.6%; Pred. No. 9.7e-56;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

```
Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 59

Qy     61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||| | |||||  ||||||||||||||||||||||||
Db     60 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 119

Qy    120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| ||||||||  |||| |||||
Db    120 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 179

Qy    178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    180 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 239

Qy    237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||||| ||||||||||||||||||
Db    240 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 273
```

RESULT 13
AAD28271
ID AAD28271 standard; DNA; 3671 BP.
XX
AC AAD28271;
XX
DT 22-APR-2002 (first entry)
XX
DE Alpha-lactalbumin (LA) Bot vector.
XX
KW Bovine; alpha-lactalbumin; promoter; pharmaceutical; industrial; ECMV;
KW encephalomyocarditis virus; diagnostic; internal ribosome entry site;
KW IRES; screening; CMV; cytomegalovirus; moloney murine leukemia virus;
KW MoMuLV; human; pre-mRNA processing enhancer; PPE; chimeric; WPRE;
KW woodchuck mRNA processing enhancer; botulinum; ds.
XX
OS Chimeric - Encephalomyocarditis virus.
OS Chimeric - Bos sp.
OS Chimeric - Homo sapiens.
OS Chimeric - Moloney murine leukemia virus.
OS Chimeric - Marmota monax.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT misc_feature 1..2053
FT /*tag= a
FT /note= "Bovine/human alpha-lactalbumin 5' flanking
FT region"
FT misc_feature 2093..2336
FT /*tag= b
FT /note= "Double mutated PPE sequence"
FT CDS 2387..2443
FT /*tag= c
FT /product= "cc49 signal peptide coding region"
FT CDS 2444..3088
FT /*tag= d
FT /product= "Bot antibody light chain Fab coding region"
FT misc_feature 3112..3686
FT /*tag= e
FT /note= "EMCV IRES"
FT CDS 3687..3745
FT /*tag= f
FT /product= "Bovine alpha-lactalbumin signal peptide
FT coding region"
FT CDS 3746..4443
FT /*tag= g
FT /product= "Bot antibody heavy chain Fab coding region"
FT misc_feature 4481..5072
FT /*tag= h
FT /note= "WPRE sequence"
FT LTR 5118..5711

```
Db    298 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 357

Qy    178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    358 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 417

Qy    237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||||| ||||||||||||||||||
Db    418 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 451


RESULT 10
AAQ54677
ID   AAQ54677 standard; DNA; 636 BP.
XX
AC   AAQ54677;
XX
DT   05-JUL-1994  (first entry)
XX
DE   Promoter/Enhancer.
XX
KW   Expression vector; promoter; enchancer; terminus; DMA1; DMA2;
KW   mutant; Namalwa cell; expression plasmid; repeat terminal;
KW   Moloney retrovirus; ss.
XX
OS   Mus musculus.
XX
PN   JP05317077-A.
XX
PD   03-DEC-1993.
XX
PF   25-MAY-1992;   92JP-0132650.
XX
PR   25-MAY-1992;   92JP-0132650.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 1994-011039/02.
XX
PT   Prepn. of protein by recombinant animal cell - using expression
PT   vector with promoter or enhancer prepd. from repeat terminal of
PT   Moloney retrovirus, and DNA for protein sequence, in Namalwa cell
XX
PS   Claim 3; page 18; 34pp; Japanese.
XX
CC   The sequence shows a promoter / enhancer DNA. The sequence can be
CC   recombined with a coding DNA within an expression plasmid which can
CC   be transformed into a Namalwa cell. This method can be used to prepare
CC   a useful protein efficiently.
XX
SQ   Sequence 636 BP; 185 A; 160 C; 152 G; 139 T; 0 other;

  Query Match           75.1%;  Score 202.8;  DB 15;  Length 636;
  Best Local Similarity   95.6%;  Pred. No. 4.7e-56;
  Matches  262;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    5;

Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    181 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 239

Qy     61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||| | |||||  |||||||||||||||||||||||||||
Db    240 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 299

Qy    120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| ||||||||  |||| |||||
Db    300 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 359

Qy    178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    360 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 419

Qy    237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||| ||||||||||||||||||
Db    420 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 453


RESULT 11
AAT32394
ID   AAT32394 standard; DNA; 702 BP.
XX
AC   AAT32394;
XX
DT   28-SEP-1996  (first entry)
XX
DE   Recombinant CMV/TAR MoMLV LTR in pMT-cat.
XX
KW   MoMLV; long terminal repeat; CMV; HIV-1; tat protein; AIDS; retrovirus;
KW   transactivation response element; TAR; vector; antiviral; virucide;
KW   antitumour; tumour; cancer; gene therapy; ds.
XX
OS   Chimeric Moloney murine leukemia virus;
OS   Chimeric human cytomegalovirus;
OS   Chimeric human immunodeficiency virus type 1.
XX
PN   WO9614332-A1.
XX
PD   17-MAY-1996.
XX
PF   08-NOV-1995;   95WO-US14576.
XX
PR   08-NOV-1994;   94US-0336132.
XX
PA   (CHAN/) CHANG L.
XX
PI   Chang L;
XX
DR   WPI; 1996-251713/25.
XX
PT   Recombinant Moloney murine leukaemia long terminal repeat which is
PT   activated by HIV-1 Tat protein  - useful in vectors for anti-viral,
PT   anti-tumour and gene therapy applications.
XX
PS   Example 3; Page 86; 126pp; English.
XX
CC   A recombinant MoMLV long terminal repeat (LTR) (AAT32394) was obtd.
CC   by replacing a portion of the U3 region of the Moloney murine
CC   leukaemia virus (MoMLV) LTR with the HIV-1 transactivation response
CC   element (TAR).  Transfection studies showed the recombinant LTR
CC   to have a low basal activity and not to be responsive to Tat
CC   protein.  This was in contrast to another recombinant LTR
CC   (AAT32382) that included the human cytomegalovirus immediate-early
CC   enhancer/promoter in addition to the TAR element.
XX
SQ   Sequence 702 BP; 170 A; 181 C; 193 G; 158 T; 0 other;

  Query Match           75.1%;  Score 202.8;  DB 17;  Length 702;
  Best Local Similarity   95.6%;  Pred. No. 4.9e-56;
  Matches  262;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    5;

Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     31 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 89

Qy     61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||| | |||||  |||||||||||||||||||||||||||
Db     90 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 149

Qy    120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| ||||||||  |||| |||||
Db    150 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 209

Qy    178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    210 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 269
```

DT 05-JUN-2000 (first entry)
XX
DE Moloney murine leukaemia virus LTR.
XX
KW Long terminal repeat; LTR; avian retrovirus;
KW constitutive transport element; CTE; replication competent;
KW host specificity; complementation; ss.
XX
OS Moloney murine leukaemia virus.
XX
PN WO200006760-A1.
XX
PD 10-FEB-2000.
XX
PF 28-JUL-1999; 99WO-US17086.
XX
PR 29-JUL-1998; 98US-0094535.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Ferris AL, Hughes SH;
XX
DR WPI; 2000-205469/18.

PT Novel avian retroviral vector with a modified constitutive transport element allowing broadened host range is useful for gene therapy -
XX
PS Claim 2; Fig 2; 36pp; English.
XX
CC The invention relates to a novel vector comprising an avian retrovirus
CC into which a nucleotide sequence comprising a portion of a non-avian
CC retroviral LTR (long terminal repeat) has been inserted, or replaces
CC all or a portion of a native constitutive transport element (CTE). The
CC vector is replication competent in at least one non-native cell type.
CC The invention also relates to a method of delivering nucleotides into a
CC target cell via the use of the novel vector; a method of making a vector
CC that is replication competent in at least one non-native cell type,
CC comprising the excision of all or a portion of the CTE of an avian-
CC derived retrovirus and replacing the CTE with a mammalian retrovirus LTR;
CC and a chicken fibroblast cell comprising a vector derived from the avian
CC leukosis virus (ALV)-based vector RCASBP which is replication competent
CC in both mammalian and chicken cells. The vector is used as a gene therapy
CC vector, or as a research tool. By altering the CTE of the virus, the host
CC range of the retrovirus can be increased. Also, there is a reduction in
CC the need for a helper virus and a reduced chance of recombination with an
CC endogenous mammalian virus. The present sequence represents the
CC Moloney murine leukaemia virus LTR which may be inserted into an avian
CC retrovirus-derived vector which has previously had all or a portion of
CC the CTE deleted.
XX
SQ Sequence 614 BP; 170 A; 164 C; 148 G; 132 T; 0 other;

Query Match 75.1%; Score 202.8; DB 21; Length 614;
Best Local Similarity 95.6%; Pred. No. 4.6e-56;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      72 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 130

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||||
Db     131 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 190

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| |||||||| |||| |||||
Db     191 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 250

Qy     178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     251 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 310

Qy     237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||| ||||||||||||||||||||
Db     311 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 344
```

RESULT 9
AAT97160
ID AAT97160 standard; DNA; 635 BP.
XX
AC AAT97160;
XX
DT 11-MAY-1998 (first entry)
XX
DE Moloney long terminal repeat.
XX
KW Vaccine; B-cell malignancy; lymphoma; leukaemia; tumour;
KW gene amplification; immunotherapy; therapy; long terminal repeat;
KW LTR; MoMLV; ds.
XX
OS Moloney murine leukemia virus.
XX
PN WO9741244-A1.
XX
PD 06-NOV-1997.
XX
PF 25-APR-1997; 97WO-US07039.
XX
PR 06-DEC-1996; 96US-0761277.
PR 01-MAY-1996; 96US-0644664.
XX
PA (GENI-) GENITOPE CORP.
XX
PI Denney DW;
XX
DR WPI; 1997-549743/50.
XX
PT Multivalent vaccine to treat B cell lymphoma or leukaemia -
PT comprises at least 2 different recombinant variable regions of
PT immunoglobulin molecules derived from B cell lymphoma cells
XX
PS Example 1; Page 104-105; 177pp; English.
XX
CC This DNA sequence comprises the Moloney murine leukaemia virus
CC (MoMLV) long terminal repeat (LTR) obtained from a plasmid
CC containing MoMLV DNA by restriction enzyme digestion. The LTR is
CC utilised in expression vectors designed for efficient expression of
CC genes in eukaryotic cells. The invention provides vectors and
CC improved methods for the expression and co-amplification of genes
CC encoding recombinant proteins in cultured cells. The methods
CC permit the isolation of cell lines which have co-amplified input
CC recombinant sequences which encode an amplifiable marker, one or
CC more expression vectors encoding a protein of interest and
CC optionally a selectable marker. The amplified cells provide large
CC quantities of recombinant proteins suitable for immunotherapy for
CC treatment of lymphomas and leukaemias. The methods permit the
CC production of custom vaccines, including multivalent vaccines that
CC reflect the degree of somatic variation found in a patient's tumour.
XX
SQ Sequence 635 BP; 184 A; 159 C; 153 G; 139 T; 0 other;

Query Match 75.1%; Score 202.8; DB 18; Length 635;
Best Local Similarity 95.6%; Pred. No. 4.7e-56;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     179 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 237

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||||
Db     238 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 297

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||||| |||||||||||||||||| |||||||| |||| |||||
```

```
PI   Von Laer M;
XX
DR   WPI; 2001-367622/38.
DR   P-PSDB; AAB86198.
XX
PT   New nucleic acid encoding membrane-anchored gp41 fusion protein, useful
PT   for gene therapy of human immunodeficiency virus (HIV) infection,
PT   prevents entry of virus into cells -
XX
PS   Claim 11; Page 28-30; 39pp; German.
XX
CC   This invention describes a novel nucleic acid (I) comprising elements
CC   that encode a signal peptide (SP) that provides transfer of expressed
CC   polypeptide into the endoplasmic reticulum, a fragment (FI) of HIV gp41
CC   protein, containing a segment from a heptad repeat region, a
CC   transmembrane anchor (MSD) of a type I membrane protein and a flexible
CC   linker (hinge) linking FI and MSD, therefore the formula of (I) is
CC   SP-FI-hinge-MSD. The products of the invention have antiviral activity
CC   and act as HIV replication inhibitors. Vectors containing (I), also T
CC   lymphocytes or hematopoietic stem cells transfected in vitro with (I),
CC   are used in gene therapy of HIV infection.
XX
SQ   Sequence 4148 BP; 880 A; 1177 C; 1103 G; 988 T; 0 other;

  Query Match             77.1%;  Score 208.2;  DB 22;  Length 4148;
  Best Local Similarity   97.1%;  Pred. No. 1.9e-57;
  Matches  265;  Conservative    0;  Mismatches    3;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||
Db    3586 GCTAGCTTAAGTAA-GCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 3643

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db    3644 AGTTCAGATCAAGGTTAGGAACAGAGAGACAGGAGAATATGGGCCAAACAGGATATCTGT 3703

Qy     121 GGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGCC 178
          |||||||||||||||||| |||||||||||||||||||||||| ||||| ||||||
Db    3704 GGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGTTGGAACAGCAGAATATGGGCC 3763

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTCC 237
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    3764 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCC 3823

Qy     238 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||||||||||||||||||||||
Db    3824 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 3856


RESULT 7
AAF83043
ID   AAF83043 standard; DNA; 508 BP.
XX
AC   AAF83043;
XX
DT   29-JUN-2001  (first entry)
XX
DE   PCR amplified MLV U3 region from pHIT111.
XX
KW   Retrovirus; recombinase recognition sequence; RRS; LTR; recombinase;
KW   long terminal repeat; pharmaceutical; cytostatic; antiinflammatory;
KW   antirheumatic; antiarthritic; antiasthmatic; osteopathic; cardiant; MLV;
KW   vasotropic; neuroprotective; nootropic; cerebroprotective; antipsoriatic;
KW   antiarteriosclerotic; vulnerary; anti-HIV; antiulcer; thrombolytic;
KW   dermatological; gene therapy; ss.
XX
OS   Murine leukemia virus.
XX
PN   WO200125466-A1.
XX
PD   12-APR-2001.
XX
PF   05-OCT-2000; 2000WO-GB03837.
XX
PR   05-OCT-1999;  99GB-0023558.
XX
PA   (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI   Slingsby J,  Kingsman SM,  Rohll J,  Slade A;
XX
DR   WPI; 2001-281732/29.
XX
PT   Modifying producer cells for making retrovirus by transfecting with a
PT   construct comprising a 5'- recombinase recognition sequence, long
PT   terminal repeat and 3'- recombinase recognition sequence, in presence
PT   of recombinase -
XX
PS   Example 7; Page 51; 133pp; English.
XX
CC   The invention relates to a method of modifying producer cells for making
CC   retrovirus by transfecting with a construct comprising a 5'- recombinase
CC   recognition sequence (RRS), long terminal repeat (LTR0 and 3'- RRS, in
CC   presence of recombinase. The regulated retroviral vector produced is
CC   useful in the manufacture of a pharmaceutical composition to deliver a
CC   NOI to a target site, and in the manufacture of a medicament for
CC   diagnostic, therapeutic and/or medical applications. The recombinase
CC   assisted method is useful for introducing regulated 3'-LTR into a derived
CC   producer cell line to produce a high titer regulated retroviral vector.
CC   The vector is useful in gene therapy for treating diseases like cancers,
CC   inflammatory diseases, immunological disorders such as graft vs host
CC   disease, autoimmune diseases such as rheumatoid arthritis, allergic
CC   diseases such as asthma, osteoporosis, cardiovascular diseases such as
CC   congestive heart failure and ischemic heart disease, neurodegenerative
CC   disorders such as multiple sclerosis, Alzheimer's disease, stroke and
CC   cerebral ischemia, atherosclerosis, thrombotic disorders, dermatological
CC   disorders such as atopic dermatitis, contact dermatitis and psoriasis,
CC   wound healing, restenosis, infectious disorders such as HIV infections,
CC   ulcers, digestive disorders such as anorexia, bulimia and cachexia, and
CC   other diseases. The present sequence represents a PCR amplified murine
CC   leukemia virus (MLV) U3 region from PHIT111, used to replace Equine
CC   infectious anemia virus (EIAV) PPT/U3 sequence. This is used in the
CC   construction of EIAV vectors with LTR driven open reading frames.
XX
SQ   Sequence 508 BP; 135 A; 129 C; 129 G; 115 T; 0 other;

  Query Match             75.1%;  Score 202.8;  DB 22;  Length 508;
  Best Local Similarity   95.6%;  Pred. No. 4.2e-56;
  Matches  262;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      58 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 116

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||||| | |||||  |||||||||||||||||||||||
Db     117 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 176

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGC 177
          |||||||||||||||||| |||||||||||||||| |||||||  |||| |||||
Db     177 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGC 236

Qy     178 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTC 236
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     237 CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC 296

Qy     237 CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||| ||||||||||||||||||
Db     297 CCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 330


RESULT 8
AAZ92704
ID   AAZ92704 standard; cDNA; 614 BP.
XX
AC   AAZ92704;
XX
```

```
XX
CC   The invention relates to a viral vector system for preparing recombinant
CC   adeno-associated virus (AAV) particles comprising; at least two plasmid
CC   vectors (ABL58983 and ABL58984) that include the two inverted terminal
CC   repeats (ITR) of AAV and additional sequences and plasmid vectors without
CC   ITRs but containing the rep and cap genes of AAV required for replication
CC   and packaging. The system is useful for producing recombinant AAV for
CC   production of a wide range of therapeutic glycoproteins in eukaryotic
CC   cells. The system provides efficient, large scale production of
CC   heterologous proteins in mammalian cells, without requiring an adenovirus
CC   helper. It is not toxic to host cells and does not cause lysis, so
CC   produced proteins are highly pure. The present sequence is that of the
CC   pAIM E1B55K plasmid of the invention.
XX
SQ   Sequence 6575 BP; 1623 A; 1566 C; 1695 G; 1691 T; 0 other;

  Query Match             91.1%;  Score 246;  DB 24;  Length 6575;
  Best Local Similarity   99.3%;  Pred. No. 1.1e-69;
  Matches  268;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||| |||||||||||| |||||||||||||||||||||||||||
Db     432 GCTAGCTTAAGTAACGCCA-TTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 489

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 549

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 609

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 669

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||||||||||||||||||
Db     670 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 699


RESULT 5
AAA96220
ID   AAA96220 standard; DNA; 9830 BP.
XX
AC   AAA96220;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Nucleotide sequence of pTRONIN, a retroviral vector.
XX
     Retroviral vector; inflammatory disorder; dermatological disorder;
     cardiovascular disorder; autoimmune disease; neurological disorder;
KW   cancer; gene therapy; ss.
XX
OS   Synthetic.
XX
PN   WO200056910-A1.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-GB01091.
XX
PR   22-MAR-1999;   99GB-0006615.
XX
PA   (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI   Uden M,  Mitrophanous K;
XX
DR   WPI; 2000-628271/60.
XX
PT   Retroviral vector for delivering one or more nucleotides of interest to
PT   a target site has functional and non-functional splice donor and splice
PT   acceptor sites -
XX
PS   Example 8; Fig 33; 148pp; English.
XX
CC   The specification describes a retroviral vector which comprises a
CC   nucleotide sequence of interest flanked by functional splice donor
CC   sites and functional splice acceptor sites. The vector is derived from
CC   a retroviral pro-vector. The retroviral vector is useful for preparing
CC   pharmaceutical compositions to deliver one or more nucleotide sequences
CC   of interest to a target sit. The retroviral vectors are especially
CC   useful for treating inflammatory disorders, cancers, dermatological
CC   disorders, cardiovascular disorders, autoimmune diseases and neurological
CC   disorders. The retroviral vector is useful in gene therapy. The present
CC   sequence represents the retroviral vector pTRONIN, a vector of the
CC   invention.
XX
SQ   Sequence 9830 BP; 2209 A; 2659 C; 2688 G; 2274 T; 0 other;

  Query Match             79.7%;  Score 215.2;  DB 21;  Length 9830;
  Best Local Similarity   94.1%;  Pred. No. 1.5e-59;
  Matches  256;  Conservative    0;  Mismatches   13;  Indels    3;  Gaps    3;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    6141 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 6199

Qy      61 AGTTCAGATCAAGGTCAGGAACAGA-GAAACAGGAGAATATGGGCCAAACAGGATATCTG 119
          |||||||||||||||||||||||||| | |||||  ||||||||||||||||||||||||
Db    6200 AGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTG 6259

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCC 178
          |||||||||||||||||||| |||||||||||||||| ||||||||  ||||  ||||
Db    6260 TGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGCC 6319

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCC 238
          |||||||||||||||||||||||||||   ||| ||||||||||||||||||||||||||
Db    6320 AAACAGGATATCTGTGGTAAGCAGTTCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCC 6379

Qy     239 CAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||| ||||||||||||||||||
Db    6380 CAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 6411


RESULT 6
AAH20890
ID   AAH20890 standard; DNA; 4148 BP.
XX
AC   AAH20890;
XX
DT   24-AUG-2001  (first entry)
XX
DE   Vector containing HIV gp41 DNA SEQ ID 1.
XX
KW   Transmembrane anchor; gene therapy; endoplasmic reticulum; gp41;
KW   antiviral; HIV replication inhibitor; T lymphocyte; viral infection;
KW   hematopoietic stem cell; ds.
XX
OS   Human immunodeficiency virus type 1.
XX
FH   Key             Location/Qualifiers
FT   CDS             1438..1773
FT                   /*tag=  a
XX
PN   WO200137881-A2.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-EP11733.
XX
PR   25-NOV-1999;   99DE-1057838.
XX
PA   (PETT-) PETTE INST HEINRICH.
XX
```

```
CC   at least two poliovirus derived internal ribosomal entry site (IRES)
CC   sequences.  The vector can be expressed in mammalian host cells for the
CC   production of heteromeric fusion proteins.  This expression system is
CC   claimed to produce the heteromeric proteins in high yields.
XX
SQ   Sequence 8298 BP; 1974 A; 2364 C; 2134 G; 1826 T; 0 other;

  Query Match             100.0%;  Score 270;  DB 19;  Length 8298;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-77;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 95

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 155

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 215

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 275

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||||||||||||||||||
Db     276 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 305


RESULT 3
AAV08560/c
ID   AAV08560 standard; cDNA; 3263 BP.
XX
AC   AAV08560;
XX
DT   12-FEB-1999  (first entry)
XX
DE   Transgene for epitope tagged TBP protein.
XX
KW   TATA-box binding protein; epitope-tagged TBP; transcription complex; TAF;
KW   TBP associated factor; TAF-interaction factor; gene expression regulator;
KW   ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   EP881288-A1.
XX
PD   02-DEC-1998.
XX
PF   26-MAY-1998;   98EP-0109516.
XX
PR   26-MAY-1997;   97EP-0108433.
XX
PA   (FARH ) HOECHST AG.
XX
PI   Berglund E,  Kirschbaum B,  Meisterernst M,  Polites G;
XX
DR   WPI; 1999-001394/01.
XX
PT   Transgenic animal expressing epitope-tagged TATA-box binding protein
PT   - for isolating higher-order transcription complexes and specific
PT   factors that associate with the protein, useful as potential
PT   therapeutic agents
XX
PS   Claim 22; Page 27-29; 38pp; English.
XX
CC   This sequence encodes an epitope-tagged TATA-box binding protein (TBP)
CC   that is expressed by the transgenic non-human animals of the invention.
CC   The animals are used to produce TBP. TBP is used to isolate and
CC   characterise higher-order transcription complexes (from different tissue
CC   and cell types, optionally at different developmental stages). It is also
CC   used to identify new and/or specific TBP associated factors (TAFs,
CC   e.g. transcription factors, activators or inhibitors) and TAF-interaction
CC   factors, and to raise antibodies against TBP. The TAFs may be useful for
CC   regulating gene expression, e.g. disease-related genes, so are potential
CC   pharmaceuticals, also for identifying human analogues for use in drug
CC   screening. The antibodies are used for affinity purification of TBP and
CC   its complexes. TBP can isolate transcription complexes from a wide
CC   variety of different tissues and cells (contrast known methods that are
CC   limited to isolation from a particular cell type).
XX
SQ   Sequence 3263 BP; 860 A; 794 C; 756 G; 853 T; 0 other;

  Query Match             94.3%;  Score 254.6;  DB 20;  Length 3263;
  Best Local Similarity   96.7%;  Pred. No. 1.2e-72;
  Matches  260;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     289 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 230

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          |||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||||
Db     229 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 170

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||| |||
Db     169 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 110

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     109 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 50

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAG 269
          |||||||||| ||||||||||||||||||
Db      49 GATGCGGTCCAGCCCTCAGCAGTTTCTAG 21


RESULT 4
ABL58984
ID   ABL58984 standard; DNA; 6575 BP.
XX
AC   ABL58984;
XX
DT   22-JUL-2002  (first entry)
XX
DE   AAV expression vector pAIM E1B55K SEQ ID NO 2.
XX
KW   Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector; ds.
XX
OS   Adeno-associated virus.
OS   Synthetic.
XX
PN   WO200238782-A2.
XX
PD   16-MAY-2002.
XX
PF   13-NOV-2001; 2001WO-EP13125.
XX
PR   13-NOV-2000; 2000DE-1056210.
XX
PA   (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.
XX
PI   Orberger G,  Hellmuth K,  Wagener C;
XX
DR   WPI; 2002-435853/46.
XX
PT   Vector system for preparing recombinant adeno-associated viral
PT   particles, used for high-level expression of heterologous therapeutic
PT   proteins in eukaryotic cells -
XX
PS   Claim 43; Page 52-56; 59pp; German.
```

```
PT   mammalian gene expression, is based on the interferon regulating
PT   factor-1 and its binding site -
XX
PS   Claim 1; Page 9; 19pp; English.
XX
CC   The present invention describes a promoter-transactivator system for
CC   inducible high-level expression of mammalian genes, and optionally
CC   for control of cell growth. The promoter-transactivator system
CC   comprises: (i) a promoter construct (IRFE promoter); and (ii)
CC   transactivator construct encoding a fusion protein of IRF-1 (interferon
CC   regulating factor-1) and the oestrogen receptor (ER). The IRFE promoter
CC   construct has the structure (MPSV-E)-(IRF-1 binding site)-(CMV)-DNA
CC   where MPSV-E indicates the myeloproliferative sarcoma virus enhancer
CC   repeat given in AAC64591; IRF-1 binding site given in AAC64592; and CMV
CC   is the cytomegalovirus minimal promoter given in AAC64593 or their
CC   functionally equivalent variants with one or more nucleotides
CC   substituted, inserted or deleted. The promoter transactivator system is
CC   a transcription regulator. Increased transcription results from binding
CC   of the IRFE promoter to the transactivator, which is activated by
CC   oestradiol or other ER ligands (these compounds displace the heat-shock
CC   protein 90 which normally binds to the IRF-1/ER fusion, preventing its
CC   activation). The system is used for increasing production of
CC   therapeutically active proteins and where a IRF-1-green fluorescent
     protein-human ER construct is used, for fluorescence-activated cell
     sorting/analysis of transformed cells.
XX
SQ   Sequence 270 BP; 85 A; 61 C; 74 G; 50 T; 0 other;

  Query Match             100.0%;  Score 270;  DB 21;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-78;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||||||||||||||||||||||
Db     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270


RESULT 2
AAV18096
ID   AAV18096 standard; DNA; 8298 BP.
XX
AC   AAV18096;
XX
DT   04-AUG-1998  (first entry)
XX
DE   pMCLDHAP tricistronic vector for the expression of hmAb45-TNF alpha.
XX
KW   Circular; antibody-cytokine fusion protein; tricistronic vector;
KW   TNF alpha; IL-2; IRES; internal ribosome entry site; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..904
FT                   /*tag= a
FT                   /note= "CMV promoter with an upstream MPSV enhancer"
FT   intron          905..976
FT                   /*tag= b
FT                   /number= Intron 1
FT   sig_peptide     977..1018
FT                   /*tag= c
FT                   /note= "Partial leader sequence"
FT   5'UTR           1019..1106
FT                   /*tag= d
FT                   /note= "5'UTR from poliovirus"
FT   CDS             1107..1433
FT                   /*tag= e
FT                   /product= "Light chain hmAb425, variable region"
FT   sig_peptide     1107..1115
FT                   /*tag= f
FT                   /note= "Rest of the leader sequence"
FT   intron          1434..1595
FT                   /*tag= g
FT                   /number= Intron 2
FT   CDS             1596..1913
FT                   /*tag= h
FT                   /product= "Light chain hmAb425, constant region"
FT   5'UTR           1914..2028
FT                   /*tag= i
FT                   /note= "5' UTR from poliovirus"
FT   RBS             2029..2159
FT                   /*tag= j
FT                   /note= "Poliovirus derived internal ribosome entry
FT                   site"
FT   intron          1260..2581
FT                   /*tag= Intron 3
FT   CDS             2582..4537
FT                   /*tag= l
FT                   /note= "Heavy chain hmAb425 fused to TNF alpha"
FT   misc_feature    4565..5279
FT                   /*tag= m
FT                   /note= "Comprises of a 5' UTR from poliovirus, an
FT                   internal ribosome entry site and intron 4"
FT   CDS             5280..5876
FT                   /*tag= n
FT                   /product= "Puromycin acetyl transferase"
FT                   /note= "Selection marker"
FT   polyA_signal    5929..6181
FT                   /*tag= o
FT                   /standard_name= SV40 PolyA
XX
PN   WO9811241-A1.
XX
PD   19-MAR-1998.
XX
PF   02-SEP-1997;   97WO-EP04765.
XX
PR   30-SEP-1996;   96EP-0115635.
PR   16-SEP-1996;   96EP-0114820.
XX
PA   (MERE ) MERCK PATENT GMBH.
XX
PI   Bruemmer W,  Burge C,  Dunker R,  Hauser H,  Mielke C;
PI   Rieke E,  Von Hoegen I,  Welge T;
XX
DR   WPI; 1998-207400/18.
DR   P-PSDB; AAW48647, AAW48648, AAW48649, AAW48650, AAW48651.
XX
PT   Oligo:cistronic expression vector - useful for production of, e.g.
PT   MAb425/TNF-^a or MAb425/IL-2 antibody fusion protein
XX
PS   Claim 11; Fig 15; 89pp; English.
XX
CC   The present sequence represents a new pMCLDHAP tricistronic vector for
CC   the expression of an antibody-cytokine fusion protein, hmAb425-TNF
CC   alpha.  hmAb425-TNF alpha comprises of the TNF alpha fused to the
CC   C-terminus of the heavy chain of humanized monoclonal antibody 425.
CC   The TNF alpha sequence can be substituted by the IL-2 sequence.  The
CC   hmAb425 has specificity for the human EGF receptor.  The vector also
CC   contains a strong promoter/enhancer unit, a selection marker gene and
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:12 ; Search time 263.478 Seconds
(without alignments)
2307.741 Million cell updates/sec

Title: US-09-980-548-1
Perfect score: 270
Sequence: 1 gctagcttaagtaacgccat.....cgccctcagcagtttctaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	270	100.0	270	21	AAC64591	Myeloproliferative
	2	270	100.0	8298	19	AAV18096	pMCLDHAP tricistro
c	3	254.6	94.3	3263	20	AAV08560	Transgene for epit
	4	246	91.1	6575	24	ABL58984	AAV expression vec
	5	215.2	79.7	9830	21	AAA96220	Nucleotide sequenc
	6	208.2	77.1	4148	22	AAH20890	Vector containing
	7	202.8	75.1	508	22	AAF83043	PCR amplified MLV
	8	202.8	75.1	614	21	AAZ92704	Moloney murine leu
	9	202.8	75.1	635	18	AAT97160	Moloney long termi
	10	202.8	75.1	636	15	AAQ54677	Promoter/Enhancer.
	11	202.8	75.1	702	17	AAT32394	Recombinant CMV/TA
	12	202.8	75.1	3097	22	AAF83096	Nucleotide sequenc
	13	202.8	75.1	3671	24	AAD28271	Alpha-lactalbumin
	14	202.8	75.1	4207	24	AAD28267	Cytomegalovirus (C
	15	202.8	75.1	4207	24	AAD28306	Cytomegalovirus (C
	16	202.8	75.1	4210	24	AAD28268	Cytomegalovirus (C
	17	202.8	75.1	4210	24	AAD28307	Cytomegalovirus (C
	18	202.8	75.1	4661	24	AAD28273	Alpha-lactalbumin
	19	202.8	75.1	4661	24	AAD28312	Alpha-lactalbumin
	20	202.8	75.1	4776	20	AAX77617	Expression constru
	21	202.8	75.1	4776	20	AAX77614	Expression constru
	22	202.8	75.1	4924	21	AAZ34937	Retrovirus vector.
	23	202.8	75.1	4924	22	AAF30946	Vector used in inv
	24	202.8	75.1	5109	18	AAT76800	pLJ Rev m10 retrov
	25	202.8	75.1	5130	24	AAD28311	LSRNL vector. Chi
	26	202.8	75.1	5176	18	AAT76801	pLJ mutant Rev m10
	27	202.8	75.1	5177	20	AAZ09997	Bovine scavenger r
	28	202.8	75.1	5594	19	AAV33629	GENSA 981, a monom
	29	202.8	75.1	5617	24	AAD32077	Human alpha-1-anti
	30	202.8	75.1	5689	20	AAZ11445	Retroviral vector
	31	202.8	75.1	5689	20	AAX61061	Retroviral vector
	32	202.8	75.1	5689	21	AAA96208	Nucleotide sequenc
	33	202.8	75.1	5691	24	AAD28274	Alpha-lactalbumin
	34	202.8	75.1	5691	24	AAD28313	Alpha-lactalbumin
	35	202.8	75.1	5711	24	AAD28310	Alpha-lactalbumin
	36	202.8	75.1	5715	21	AAZ34936	Retrovirus vector.
	37	202.8	75.1	5715	22	AAF30945	Vector used in inv
	38	202.8	75.1	5731	24	AAD32078	Human alpha-1-anti
	39	202.8	75.1	5732	24	AAD28269	Mouse mammary tumo
	40	202.8	75.1	5732	24	AAD28308	Mouse mammary tumo
	41	202.8	75.1	5865	19	AAV04002	Retroviral vector
	42	202.8	75.1	5874	20	AAX90484	Plasmid retroviral
	43	202.8	75.1	6026	24	AAD32075	Human albumin prom
	44	202.8	75.1	6140	24	AAD32076	Human albumin prom
	45	202.8	75.1	6141	20	AAX90481	Plasmid retroviral

ALIGNMENTS

RESULT 1
AAC64591
ID AAC64591 standard; DNA; 270 BP.
XX
AC AAC64591;
XX
DT 15-FEB-2001 (first entry)
XX
DE Myeloproliferative sarcoma virus enhancer sequence SEQ ID NO:1.
XX
KW Myeloproliferative sarcoma virus; MPSV; MPSV-E; enhancer; CMV;
KW cytomegalovirus; firefly; IRF-1 binding site; minimal promoter;
KW interferon regulatory factor 1 binding site; transcription regulator;
KW promoter transactivator system; ds.
XX
OS Myeloproliferative sarcoma virus.
XX
PN EP1046710-A1.
XX
PD 25-OCT-2000.
XX
PF 23-APR-1999; 99EP-0108068.
XX
PR 23-APR-1999; 99EP-0108068.
XX
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI Mueller PP, Geserick C, Schroeder K, Hauser H;
XX
DR WPI; 2000-648930/63.
XX
PT Promoter-transactivator system, useful for inducing high level

```
                  /bound_moiety="glucocorticoid receptor"
     protein_bind 263. .275
                  /bound_moiety="nuclear factor (NF)"
     enhancer     287. .294
     repeat_region 320. .392
                  /rpt_type=direct
     protein_bind 321. .326
                  /bound_moiety="glucocorticoid receptor"
     protein_bind 338. .349
                  /bound_moiety="nuclear factor (NF)"
     enhancer     361. .368
     protein_bind 396. .400
                  /bound_moiety="glucocorticoid receptor"
     CAAT_signal  508. .512
     TATA_signal  559. .565
     mRNA         590. .733
                  /partial
     repeat_region 590. .657
                  /note="R region"
     repeat_region 658. .733
                  /note="U5 region"
     repeat_region 721. .733
                  /rpt_type=inverted
BASE COUNT      195 a    193 c    176 g    168 t      1 others
ORIGIN

  Query Match             76.4%;  Score 206.4;  DB 14;  Length 733;
  Best Local Similarity   94.1%;  Pred. No. 1.6e-53;
  Matches  257;  Conservative    0;  Mismatches   12;  Indels    4;  Gaps    4;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     170 GCTAGCTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 228

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAG-AAACAGGAGAATATGGGCCAAACAGGATATCTG 119
           |||||||||||||||||||||||||  ||||||  |||||||||||| ||||||||||||
Db     229 AGTTCAGATCAAGGTCAGGAACAGANAAAACAGCTGAATATGGGCCCAACAGGATATCTG 288

Qy     120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCC 178
           ||| ||||||||||||||| |||||||||||||||| || | |||  ||||||||||
Db     289 TGGCAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGAATCAGCTGAATTGGGCC 348

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTCC 237
           |||||||||||||| |||||||||||||||||| |||||||||||||||||||||||||
Db     349 AAACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCC 408

 y     238 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||||||| |||||||||||||||||||
Db     409 CCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGA 441


Search completed: March  5, 2003, 23:28:17
Job time : 1503.61 secs
```

```
                     SignalPEPTIDEkodiert
                     T-20-Fusionsprotein mit Membrananker"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC42734.1"
                     /db_xref="GI:14546735"
                     /translation="MGAGATGRAMDGPRLLLLLLLGVSLGGARSYTSLIHSLIEESQN
                     QQEKNEQELLELDKWASLWNWFRSVPRDCGCKPCICTLIPVYCSILAAVVVGLVAYIA
                     FKRWNRGIL"
     sig_peptide     1438. .1527
     misc_feature    1438. .1773
                     /note="STHM (Vektor-Insert, das den fuer das
                     Membran-verankerte T-20-PEPTIDEkodierenden Bereich
                     enthaelt)"
     misc_feature    1528. .1635
                     /note="Fuer T-20 kodierender Bereich"
     misc_feature    1636. .1683
                     /note="Fuer Hinge kodierender Bereich"
     misc_feature    1684. .1773
                     /note="Fuer Membrananker kodierender Bereich"
     misc_feature    1775. .1780
                     /note="SalI-Schnittstelle"
     misc_feature    1787. .1792
                     /note="SphI-Schnittstelle"
     misc_feature    1793. .2421
                     /note="Polio-IRES"
     misc_feature    1793. .1798
                     /note="HindIII-Schnittstelle"
     misc_feature    2380. .3267
                     /note="neoR cDNA"
     misc_feature    2422. .2429
                     /note="NotI-Schnittstelle"
     misc_feature    3002. .3007
                     /note="SphI-Schnittstelle"
     misc_feature    3447. .3452
                     /note="HindIII-Schnittstelle"
     misc_feature    3586. .3591
                     /note="NheI-Schnittstelle"
     misc_feature    4000. .4007
                     /note="AscI-Schnittstelle"
     misc_feature    4030. .4035
                     /note="XmaI-Schnittstelle"
BASE COUNT      880 a   1177 c   1103 g    988 t
ORIGIN

  Query Match             77.1%;  Score 208.2;  DB 6;  Length 4148;
  Best Local Similarity   97.1%;  Pred. No. 5.5e-54;
  Matches  265;  Conservative    0;  Mismatches    3;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           ||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||
Db    3586 GCTAGCTTAAGTAA-GCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 3643

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           ||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db    3644 AGTTCAGATCAAGGTTAGGAACAGAGAGACAGGAGAATATGGGCCAAACAGGATATCTGT 3703

Qy     121 GGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGCC 178
           |||||||||||||||||| |||||||||||||||||||||||||| ||||| ||||||
Db    3704 GGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGTTGGAACAGCAGAATATGGGCC 3763

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTCC 237
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    3764 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCC 3823

Qy     238 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||||||||||||||||||||||||||
Db    3824 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 3856


RESULT 14
AR028672
LOCUS       AR028672               9318 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 6 from patent US 5858744.
ACCESSION   AR028672
VERSION     AR028672.1  GI:5940645
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 9318)
  AUTHORS   Baum,C., Stocking-Harbers,C. and Ostertag,W.
  TITLE     Retroviral vector hybrids and the use thereof for gene transfer
  JOURNAL   Patent: US 5858744-A 6 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .9318
                     /organism="unknown"
BASE COUNT     2377 a   2217 c   2364 g   2358 t      2 others
ORIGIN

  Query Match             77.1%;  Score 208.2;  DB 6;  Length 9318;
  Best Local Similarity   97.1%;  Pred. No. 6e-54;
  Matches  265;  Conservative    0;  Mismatches    3;  Indels    5;  Gaps    5;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           ||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||
Db    5899 GCTAGCTTAAGTAA-GCCATTTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 5956

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           ||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db    5957 AGTTCAGATCAAGGTTAGGAACAGAGAGACAGGAGAATATGGGCCAAACAGGATATCTGT 6016

Qy     121 GGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGCC 178
           |||||||||||||||||| |||||||||||||||||||||||||| ||||| ||||||
Db    6017 GGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGTTGGAACAGCAGAATATGGGCC 6076

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTCC 237
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    6077 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCC 6136

Qy     238 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||||||||||||||||||||||||||
Db    6137 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 6169


RESULT 15
MLM3LTR
LOCUS       MLM3LTR                733 bp ss-RNA     linear   VRL 09-DEC-1994
DEFINITION  Moloney murine leukemia virus TB 3' LTR region.
ACCESSION   M24204
VERSION     M24204.1  GI:600759
KEYWORDS    LTR.
SOURCE      Moloney murine leukemia virus (strain TB) RNA.
  ORGANISM  Moloney murine leukemia virus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   Yuen,P.H. and Szurek,P.F.
  TITLE     The reduced virulence of the thymotropic Moloney murine leukemia
            virus derivative MoMuLV-TB is mapped to 11 mutations within the U3
            region of the long terminal repeat
  JOURNAL   J. Virol. 63 (2), 471-480 (1989)
  MEDLINE   89094971
   PUBMED   2783465
COMMENT     On Dec 10, 1994 this sequence version replaced gi:341122.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Moloney murine leukemia virus"
                     /strain="TB"
                     /db_xref="taxon:11801"
     repeat_region   140. .590
                     /note="U3 region"
     repeat_region   140. .153
                     /rpt_type=inverted
     repeat_region   245. .319
                     /rpt_type=direct
     protein_bind    246. .251
```

```
FEATURES             Location/Qualifiers
     source          1. .2854
                     /organism="Myeloproliferative sarcoma virus"
                     /db_xref="taxon:11813"
     CDS             973. .2001
                     /note="v-mos protein"
                     /codon_start=1
                     /protein_id="AAA46580.1"
                     /db_xref="GI:332209"
                     /translation="MPSPLSLCRYLPRELSPSVDSRSCSIPLVAPRKAGKLFLGTTPP
                     RAPGLPRRLAWFSIDWEQVCLMHRLGSGGFGSVYKATYHGVPVAIKQVNKCTKDLRAS
                     QRSFWAELNIARLRHDNIVRVVAASTRTPEDSNSLGTIIMEFGGNVTLHQVIYGATRS
                     PEPLSCREQLSLGKCLKYSLDVVNGLLFLHSQSILHLDLKPANILISEQDVCKISDFG
                     CSQKLQDLRCRQSPHHIGGTYTHQAPEILKGEIATPKADIYSFGITLWQMTTREVPYS
                     GEPQYVQYAVVAYNLRPSLTGAVFTASLTGKTLQNIIQNCWEARALQRPGAELLQRDL
                     KAFRGALG"
     LTR             2268. .2854
                     /note="3' long terminal repeat"
     repeat_region   2383. .2456
                     /note="tandem repeat A"
     repeat_region   2457. .2530
                     /note="tandem repeat B"
BASE COUNT      696 a    814 c    709 g    635 t
ORIGIN      1 bp upstream of HindIII site.

  Query Match             82.6%;  Score 223;  DB 14;  Length 2854;
  Best Local Similarity   98.5%;  Pred. No. 1.3e-58;
  Matches  267;  Conservative    0;  Mismatches    0;  Indels    4;  Gaps    4;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db    2298 GCTAGCTTAAGTAACGCCA-TTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 2355

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2356 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 2415

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGCCA 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    2416 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATATGGGCCA 2475

Qy     180 AACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCC 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2476 AACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCC 2535

Qy     240 AGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||| |||||||||||||||||||||||||||
Db    2536 AGAT-CGGTCCCGCCCTCAGCAGTTTCTAGA 2565


RESULT 12
S66424/c
LOCUS       S66424                   544 bp    DNA     linear   ROD 06-JAN-1994
DEFINITION  c-myc {5' LTR, provirus insertion} [mice, plasmacytoma RFPC 2782,
            Genomic Mutant, 544 nt}.
ACCESSION   S66424
VERSION     S66424.1  GI:439771
KEYWORDS    .
SOURCE      Mus sp. plasmacytoma RFPC 2782.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Shaughnessy,J.D. Jr., Owens,J.D. Jr., Wiener,F., Hilbert,D.M.,
            Huppi,K., Potter,M. and Mushinski,J.F.
  TITLE     Retroviral enhancer insertion 5' of c-myc in two
            translocation-negative mouse plasmacytomas upregulates c-myc
            expression to different extents
  JOURNAL   Oncogene 8 (11), 3111-3121 (1993)
  MEDLINE   94020837
   PUBMED   8414513
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 138854] from the original journal article.

            This sequence comes from Fig. 4b.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            1. .544
                     /gene="c-myc"
BASE COUNT      121 a    129 c    117 g    177 t
ORIGIN

  Query Match             77.4%;  Score 209;  DB 10;  Length 544;
  Best Local Similarity   94.9%;  Pred. No. 2.4e-54;
  Matches  259;  Conservative    0;  Mismatches   10;  Indels    4;  Gaps    4;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||  |||||||||||||||||||||||||| ||||||||| ||||||||||||||||||
Db     392 GCTACGTTAAGTAACGCCATTTTGCAAGGCAT-GGAAAAATAGATAACTGAGAATAGAGA 334

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db     333 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGCTGAATATGGGCCAAACAGGATATCTGT 274

Qy     121 GGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGTTGGAACAGGAGAAT-TGGGCC 178
          |||||||||||||||||| |||||||||||||||| ||||||| | |||| ||||||
Db     273 GGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACACGTGAATATGGGCC 214

Qy     179 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAAGAACAGATGGTCC 237
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     213 AAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCC 154

Qy     238 CCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||| ||||||||||||||||||
Db     153 CCAGATGCGGTCCACCCCTCAGCAGTTTCTAGA 121


RESULT 13
AX166278
LOCUS       AX166278                 4148 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 1 from Patent WO0137881.
ACCESSION   AX166278
VERSION     AX166278.1  GI:14546734
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 4148)
  AUTHORS   von Laer,M.D.
  TITLE     Gene therapy of hiv-positive patients by the expression of
            membrane-anchored gp41 peptides
  JOURNAL   Patent: WO 0137881-A 1 31-MAY-2001;
            Heinrich-Pette-Institut (DE)
FEATURES             Location/Qualifiers
     source          1. .4148
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="M87 (STHM) - Retroviraler Vektor MPIN, der 5' von
                     der IRES-NEO-Kassette ein Insert enthaelt, das fuer
                     Membran-verankertes T-20-PEPTIDEkodiert"
     misc_feature    31. .36
                     /note="NheI-Schnittstelle"
     misc_feature    370. .377
                     /note="AscI-Schnittstelle"
     misc_feature    400. .405
                     /note="XmaI-Schnittstelle"
     misc_feature    696. .701
                     /note="SacII-Schnittstelle"
     misc_feature    1421. .1426
                     /note="SacII-Schnittstelle"
     misc_feature    1429. .1436
                     /note="NotI-Schnittstelle"
     CDS             1438. .1773
                     /note="Bereich im M87 (STHM)-Insert, der fuer
                     Membran-verankertes T-20-Fusionsprotein einschliesslich
```

```
                promoter is 993"
                /citation=[3]
     misc_feature    1061. .1076
                /note="multiple cloning site (MCS); contains the unique
                restriction sites XhoI, SacII, NotI"
     intron          1160. .1224
                /note="SV40 small T antigen intron provides splice signal
                for inserts cloned into the MCS"
     polyA_signal    1860. .1865
                /note="The polyA signal from the SV40 early region
                provides polyadenylation of inserts cloned in the MCS"
     polyA_site      1881
     misc_feature    1931. .9875
                /note="complete genome of bovine papilloma virus type 1;
                The BPV genome provides for episomal maintenance of the
                vector pBPV when the vector is transfected into mammalian
                cells"
                /citation=[5]
     rep_origin      complement(10647)
                /note="Plasmid origin of replication for replication in E.
                coli. base 10647 represents the first base of the newly
                synthesized strand"
                /direction=left
     gene            complement(11406. .12336)
                /gene="bla"
     CDS             complement(11406. .12266)
                /gene="bla"
                /codon_start=1
                /transl_table=11
                /product="beta-lactamase"
                /protein_id="AAA57075.1"
                /db_xref="GI:595689"
                /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVE
                YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                DRWEPELNEAIPNDERDTTMPAAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                EIGASLIKHW"
     promoter        complement(12308. .12336)
                /gene="bla"
BASE COUNT     3381 a    2852 c    3005 g    3278 t
ORIGIN

  Query Match            94.3%;  Score 254.6;  DB 12;  Length 12516;
  Best Local Similarity  96.7%;  Pred. No. 2e-68;
  Matches  260;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 210

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          |||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||||
Db     209 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 150

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          |||||||||||||||||||||||||||||||||||| ||||||||  |||||||| |||
Db     149 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 90

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 30

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAG 269
          |||||||||| ||||||||||||||||||
Db      29 GATGCGGTCCAGCCCTCAGCAGTTTCTAG 1


RESULT 10
AX449148
LOCUS       AX449148               6575 bp    DNA     linear   PAT 03-JUL-2002
DEFINITION  Sequence 2 from Patent WO0238782.
ACCESSION   AX449148
VERSION     AX449148.1  GI:21697950
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Orberger,G., Hellmuth,K. and Wagener,C.
  TITLE     Viral expression system
  JOURNAL   Patent: WO 0238782-A 2 16-MAY-2002;
            Arimedes Biotechnology GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .6575
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Plasmid pAIM E1B55K, das eine Expressionskassette
                     fuer E1B55K enth 1t."
BASE COUNT     1623 a    1566 c    1695 g    1691 t
ORIGIN

  Query Match            91.1%;  Score 246;  DB 6;  Length 6575;
  Best Local Similarity  99.3%;  Pred. No. 9.1e-66;
  Matches  268;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          |||||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db     432 GCTAGCTTAAGTAACGCCA-TTTGCAAGGCAT-GGAAAAATACATAACTGAGAATAGAGA 489

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 549

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 609

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 669

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||||||||||||||||||
Db     670 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 699


RESULT 11
MSYMOS
LOCUS       MSYMOS                  2854 bp ss-RNA     linear   VRL 02-AUG-1993
DEFINITION  Myeloproliferative sarcoma virus proviral v-mos gene, 3' LTR.
ACCESSION   K01683
VERSION     K01683.1  GI:332208
KEYWORDS    c-myc proto-oncogene; coat protein; envelope protein;
            envelope-associated protein; mos oncogene.
SOURCE      Myeloproliferative sarcoma virus proviral DNA.
  ORGANISM  Myeloproliferative sarcoma virus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 2854)
  AUTHORS   Stacey,A., Arbuthnott,C., Kollek,R., Coggins,L. and Ostertag,W.
  TITLE     Comparison of myeloproliferative sarcoma virus with Moloney murine
            sarcoma virus variants by nucleotide sequencing and heteroduplex
            analysis
  JOURNAL   J. Virol. 50 (3), 725-732 (1984)
  MEDLINE   84216451
   PUBMED   6328002
COMMENT     [1] reports an amber mutation (deletion of one 'c') in the envelope
            gene of the MPSV genome, at the junction of the envelope/mos genes.
            What would normally be the twentieth codon is now a termination
            codon. At base 973 an in-phase start codon was found. The MPSV
            sequence is compared with two temperature sensitive mutants derived
            from it, Mo-MuSV variant M1 and Moloney murine leukemia virus
            (Mo-MuLV) by heteroduplex mapping. MPSV wild-type contained a 1 kb
            deletion from the pol gene. It also contained sequences related to
            v-mos genes.
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 230

Qy     61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
                ||||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||
Db    229 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 170

Qy    121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
                |||||||||||||||||||||||||||||||||||| ||||||||  |||||||| |||
Db    169 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 110

Qy    181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
                ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    109 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 50

Qy    241 GATGCGGTCCCGCCCTCAGCAGTTTCTAG 269
                |||||||||| ||||||||||||||||||
      49 GATGCGGTCCAGCCCTCAGCAGTTTCTAG 21


RESULT 8
E28258/c
LOCUS       E28258                  3263 bp    DNA     linear   PAT 18-JUN-2001
DEFINITION  Purification of higher transcription complex from nonhuman
            transgenic animal.
ACCESSION   E28258
VERSION     E28258.1  GI:13025292
KEYWORDS    JP 1999004638-A/10.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3263)
  AUTHORS   Bernd,K., Eric,B., Michael,M. and Greg,P.
  TITLE     Purification of higher transcription complex from nonhuman
            transgenic animal
  JOURNAL   Patent: JP 1999004638-A 10 12-JAN-1999;
            HOECHST AG
COMMENT     OS   Unidentified
            PN   JP 1999004638-A/10
            PD   12-JAN-1999
            PF   26-MAY-1998 JP 1998144743
            PR   26-MAY-1997 DE
            PI   BERND KIRSCHBAUM,ERIC BERUGURUNTO,MICHAEL MEISTERERNST, PI
            GREG PORITTSU
            PC   A01K67/027,C12N15/09,C12P21/02//C12P21/08,(C12N15/09,
            C12R1:91), PC   C12N15/00,
            PC   (C12N15/00,C12R1:91)
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   exon            1. .3263.
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      860 a    794 c    756 g    853 t
ORIGIN

  Query Match             94.3%;  Score 254.6;  DB 6;  Length 3263;
  Best Local Similarity   96.7%;  Pred. No. 1.7e-68;
  Matches  260;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 230

Qy     61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
                ||||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||
Db    229 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 170

Qy    121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
                |||||||||||||||||||||||||||||||||||| ||||||||  |||||||| |||
Db    169 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 110

Qy    181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
                ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    109 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 50

Qy    241 GATGCGGTCCCGCCCTCAGCAGTTTCTAG 269
                |||||||||| ||||||||||||||||||
      49 GATGCGGTCCAGCCCTCAGCAGTTTCTAG 21


RESULT 9
XXU13843/c
LOCUS       XXU13843                12516 bp    DNA     circular SYN 13-DEC-1994
DEFINITION  pBPV cloning vector, complete sequence.
ACCESSION   U13843
VERSION     U13843.1  GI:595688
KEYWORDS    bovine papilloma virus; metallothionein I promoter; Moloney sarcoma
            virus enhancer; beta-lactamase.
SOURCE      unidentified cloning vector.
  ORGANISM  unidentified cloning vector
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 12516)
  AUTHORS   Malone,J.A.
  TITLE     pBPV: An episomally maintained mammalian expression vector based on
            bovine papilloma virus
  JOURNAL   Unpublished (1994)
REFERENCE   2  (bases 1 to 391)
  AUTHORS   Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
  TITLE     Nucleotide sequences of integrated Moloney sarcoma provirus long
            terminal repeats and their host and viral junctions
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
  MEDLINE   81054687
   PUBMED   6254003
REFERENCE   3  (bases 692 to 1060)
  AUTHORS   Glanville,N., Durnam,D.M. and Palmiter,R.D.
  TITLE     Structure of mouse metallothionein-I gene and its mRNA
  JOURNAL   Nature 292 (5820), 267-269 (1981)
  MEDLINE   81245168
   PUBMED   7254320
REFERENCE   4  (bases 939 to 949)
  AUTHORS   Searle,P.F., Stuart,G.W. and Palmiter,R.D.
  TITLE     Building a metal-responsive promoter with synthetic regulatory
            elements
  JOURNAL   Mol. and Cell. Biol. 5, 1480-1489 (1985)
REFERENCE   5  (bases 1931 to 9875)
  AUTHORS   Chen,E.Y., Howley,P.M., Levinson,A.D. and Seeburg,P.H.
  TITLE     The primary structure and genetic organization of the bovine
            papillomavirus type 1 genome
  JOURNAL   Nature 299 (5883), 529-534 (1982)
  MEDLINE   83012974
   PUBMED   6289124
REFERENCE   6  (bases 1 to 12516)
  AUTHORS   Malone,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-1994) James A. Malone, International Technical
            Services, Molecular Biology Reagents Division, Pharmacia Biotech
            Inc., 2202 N. Bartlett Ave., Milwaukee, WI 53202-1009, USA
FEATURES             Location/Qualifiers
     source          1. .12516
                     /organism="unidentified cloning vector"
                     /db_xref="taxon:45196"
                     /lab_host="Escherichia coli"
     enhancer        complement(1. .391)
                     /standard_name="Moloney Sarcoma Virus (MSV) enhancer"
                     /citation=[2]
     misc_signal     939. .949
                     /standard_name="metal response element (MRE)"
                     /citation=[4]
     gene            965. .970
                     /gene="mouse metallothionein I"
     TATA_signal     965. .970
                     /gene="mouse metallothionein I"
                     /note="The first base of nascent transcripts from the MT-I
```

```
SOURCE      MMSV (Moloney murine sarcoma virus) from transformed mink cells,
            clone P-600-3.
  ORGANISM  Moloney murine leukemia virus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
  TITLE     Nucleotide sequences of integrated Moloney sarcoma provirus long
            terminal repeats and their host and viral junctions
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
  MEDLINE   81054687
   PUBMED   6254003
COMMENT     On Oct 4, 1994 this sequence version replaced gi:331983.
            [1]: the first and last 11 nucleotides of the terminal repeat
            sequence (trs) are inverted with respect to each other, and the
            same four-nucleotide host sequence is found bracketing integrated
            MSV. Bases 233-820 represent the 588 bp left trs. See mmsvtrs2 and
            mmsvtrs3 for the following segments. See also mmsvpro for the
            complete proviral sequence.
               EMBL features not translated to GenBank features:
               key         from     to       description
               CELL          <1     232      cellular DNA (from mink)
               PROVRL       233    >903      proviral DNA
               SITE         233     820      terminal repeat
               SITE         229     232      repeated again at right end
                                             of provirus.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Moloney murine leukemia virus"
                     /proviral
                     /db_xref="taxon:11801"
BASE COUNT      238 a    245 c    215 g    205 t
ORIGIN      352 bases upstream of PvuII site.

  Query Match             94.7%;  Score 255.6;  DB 14;  Length 903;
  Best Local Similarity   96.7%;  Pred. No. 7.2e-69;
  Matches  261;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 320

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           |||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||
Db     321 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 380

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
           |||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db     381 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 440

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
       441 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 500

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||| ||||||||||||||||||||
Db     501 GATGCGGTCCAGCCCTCAGCAGTTTCTAGA 530


RESULT 6
REMSV3
LOCUS       REMSV3                   903 bp    DNA     linear   VRL 04-APR-2000
DEFINITION  Left long terminal repeat of Moloney Murine sarcoma virus. (as
            integrated into mink DNA).
ACCESSION   V01182 J02268
VERSION     V01182.1  GI:61640
KEYWORDS    terminal repeat.
SOURCE      Moloney murine sarcoma virus.
  ORGANISM  Moloney murine sarcoma virus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
  TITLE     Nucleotide sequences of integrated Moloney sarcoma provirus long
            terminal repeats and their host and viral junctions
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
  MEDLINE   81054687
   PUBMED   6254003
FEATURES             Location/Qualifiers
     source          233. .903
                     /organism="Moloney murine sarcoma virus"
                     /proviral
                     /db_xref="taxon:11809"
                     /focus
     source          1. .232
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
     misc_feature    229. .232
                     /note="repeated again at right end of provirus"
     misc_feature    233. .820
                     /note="terminal repeat"
BASE COUNT      238 a    245 c    215 g    205 t
ORIGIN

  Query Match             94.7%;  Score 255.6;  DB 14;  Length 903;
  Best Local Similarity   96.7%;  Pred. No. 7.2e-69;
  Matches  261;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 320

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
           |||||||||||||||||||||||||| ||||  ||||||||||||||||||||||||||
Db     321 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 380

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
           |||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db     381 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 440

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     441 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 500

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
           |||||||||| ||||||||||||||||||||
Db     501 GATGCGGTCCAGCCCTCAGCAGTTTCTAGA 530


RESULT 7
AX002278/c
LOCUS       AX002278                 3263 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 15 from Patent EP0881288.
ACCESSION   AX002278
VERSION     AX002278.1  GI:7241968
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3263)
  AUTHORS   Berglund,E.D. and Kirschbaum,B.D.
  TITLE     Purification of higher order transcription complexes from
            transgenic non-human animals
  JOURNAL   Patent: EP 0881288-A 15 02-DEC-1998;
            HOECHST AG (DE)
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     exon            1. .3263
BASE COUNT      860 a    794 c    756 g    853 t
ORIGIN

  Query Match             94.3%;  Score 254.6;  DB 6;  Length 3263;
  Best Local Similarity   96.7%;  Pred. No. 1.7e-68;
  Matches  260;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
```

```
                 APKLLIYDTSNLASGVPSRFSGSGSGTDYTFTISSLQPEDIATYYCQQWSSHIFTFGQ
                 GTKVEIK"
     intron      1434. .1595
     CDS         <1596. .1916
                 /codon_start=1
                 /transl_table=11
                 /product="LIGHT CHAIN HMAB425, CONSTANT REGION"
                 /protein_id="CAB42568.1"
                 /db_xref="GI:4774644"
                 /translation="TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDN
                 ALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFN
                 RGEC"
     CDS         2582. .4540
                 /codon_start=1
                 /transl_table=11
                 /product="FUSION PROTEIN: HEAVY CHAIN HMAB425 +
                 TNFALPHA(FROM 4064)"
                 /protein_id="CAB42569.1"
                 /db_xref="GI:4774645"
                 /translation="MDWTWRVFCLLAVAPGAHSQVQLVQSGAEVKKPGASVKVSCKAS
                 GYTFTSHWMHWVRQAPGQGLEWIGEFNPSNGRTNYNEKFKSKATMTVDTSTNTAYMEL
                 SSLRSEDTAVYYCASRDYDYDGRYFDYWGQGTLVTVSSGEWILCAWAQLCPTPRSHGT
                 TSLAASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFP
                 AVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPC
                 PAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNA
                 KTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPRE
                 PQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGS
                 FFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKMVRSSSRTPSDKPV
                 AHVVANPQAEGQLQWLNRRANALLANGVELRDNQLVVPSEGLYLIYSQVLFKGQGCPS
                 THVLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKG
                 DRLSAEINRPDYLDFAESGQVYFGIIAL"
     misc_RNA    4565. .5279
                 /product="5'UTR FROM POLIVIRUS PLUS IRES PLUS INTRON"
     CDS         5280. .5879
                 /function="SELECTION MARKER"
                 /codon_start=1
                 /transl_table=11
                 /product="PUROMYCIN ACETYL TRANSFERASE"
                 /protein_id="CAB42570.1"
                 /db_xref="GI:4774646"
                 /translation="MTEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIE
                 RVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPESVEAGAVFAEIGPRMAELSGSR
                 LAAQQQMEGLLAPHRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAAERAGVPAFL
                 ETSAPRNLPFYERLGFTVTADVECPKDRATWCMTRKPGA"
     misc_RNA    5877. .8298
                 /product="DNA SEQUENCE COMPRISING SV40 POLYA (5929-6181)"
                 /standard_name="SV40 POLYA"
BASE COUNT     1974 a    2364 c    2134 g    1826 t
ORIGIN
```

Query Match 100.0%; Score 270; DB 6; Length 8298;
Best Local Similarity 100.0%; Pred. No. 3e-73;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 95

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 155

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 215

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 275

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||||||||||||||||||||||
Db     276 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 305
```

```
RESULT 4
REMSV4
LOCUS       REMSV4                    646 bp    DNA     linear   VRL 04-APR-2000
DEFINITION  Moloney mouse sarcoma virus right long terminal repeat (as
            integrated into mink DNA).
ACCESSION   V01183
VERSION     V01183.1  GI:61641
KEYWORDS    terminal repeat.
SOURCE      Moloney murine sarcoma virus.
  ORGANISM  Moloney murine sarcoma virus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
  TITLE     Nucleotide sequences of integrated Moloney sarcoma provirus long
            terminal repeats and their host and viral junctions
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
  MEDLINE   81054687
   PUBMED   6254003
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Moloney murine sarcoma virus"
                     /db_xref="taxon:11809"
                     /focus
                     /note="proviral DNA integrated into Mink"
     source          1. .646
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
     source          595. .646
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /note="cellular DNA"
     misc_feature    7. .594
                     /note="right terminal repeat"
     misc_feature    595. .598
                     /note="repeated before left end of provirus"
BASE COUNT      162 a    174 c    164 g    146 t
ORIGIN
```

Query Match 94.7%; Score 255.6; DB 14; Length 646;
Best Local Similarity 96.7%; Pred. No. 6.9e-69;
Matches 261; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 94

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          |||||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||
Db      95 AGTTCAGATCAAGGTCAGGAACAGAGAGACAGCTGAATATGGGCCAAACAGGATATCTGT 154

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          |||||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db     155 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGAACAGCTGAATTGGGGCAA 214

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 ACAGGATATCTGCGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 274

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          |||||||||| ||||||||||||||||||||
Db     275 GATGCGGTCCAGCCCTCAGCAGTTTCTAGA 304
```

```
RESULT 5
MLMTRS01
LOCUS       MLMTRS01                  903 bp ss-RNA     linear   VRL 02-AUG-1993
DEFINITION  Moloney murine sarcoma provirus LTR and host sequence, segment 1.
ACCESSION   J02268
VERSION     J02268.1  GI:555246
KEYWORDS    terminal repeat.
SEGMENT     1 of 3
```

```
  JOURNAL   Patent: WO 0065074-A 1 02-NOV-2000;
            Gesellschaft f}r Biotechnologische Forschung mbH (GBF) (; DE)
FEATURES             Location/Qualifiers
     source          1. .270
                     /organism="Myeloproliferative sarcoma virus"
                     /db_xref="taxon:11813"
BASE COUNT       85 a     61 c     74 g     50 t
ORIGIN

  Query Match             100.0%;  Score 270;  DB 6;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 2e-73;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||||||||||||||||||||
Db     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270


RESULT 2
PMBC2TD
LOCUS       PMBC2TD                 3123 bp    DNA     linear   SYN 23-JAN-1995
DEFINITION  Plasmid pMBC-2T (expression plasmid-eukaryotic-in vitro) DNA.
ACCESSION   X77750
VERSION     X77750.1  GI:456203
KEYWORDS    expression plasmid; Plasmid.
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 3123)
  AUTHORS   Dirks,W., Schaper,F. and Hauser,H.
  TITLE     A new hybrid promoter directs transcription at identical starts
            points in mammalian cells and in vitro
  JOURNAL   Gene 49, 389-390 (1994)
REFERENCE   2  (bases 1 to 3123)
  AUTHORS   Schaper,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1994) F. Schaper, Gesellschaft fr.
            Biotechnologische Fors., (GBF), Mascheroder Weg 1, 38106
            Braunschweig, FRG
FEATURES             Location/Qualifiers
     source          1. .3123
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /plasmid="pMBC-2T"
     promoter        6. .455
                     /note="from MT7"
     gene            463. .521
                     /gene="SV40 19S mRNA"
     intron          463. .521
                     /gene="SV40 19S mRNA"
                     /note="from VP1"
     misc_feature    529. .591
                     /note="polylinker site from pUC8"
     polyA_signal    626. .862
                     /note="SV40 of VP1"
     terminator      1022. .1157
                     /note="from T7"
BASE COUNT      815 a    753 c    761 g    794 t
ORIGIN

  Query Match             100.0%;  Score 270;  DB 12;  Length 3123;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-73;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36 GCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGGAAAAATACATAACTGAGAATAGAGA 95

Qy      61 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 AGTTCAGATCAAGGTCAGGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGT 155

Qy     121 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 GGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATTGGGCCAA 215

Qy     181 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 ACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAAGAACAGATGGTCCCCA 275

Qy     241 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 270
          ||||||||||||||||||||||||||||||
Db     276 GATGCGGTCCCGCCCTCAGCAGTTTCTAGA 305


RESULT 3
A70359
LOCUS       A70359                  8298 bp    DNA     circular PAT 07-MAY-1999
DEFINITION  Sequence 1 from Patent WO9811241.
ACCESSION   A70359
VERSION     A70359.1  GI:4774641
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 8298)
  AUTHORS   Von,H.I., Bruemmer,W., Burge,C., Rieke,E., Dunker,R., Welge,T.,
            Hauser,H. and Mielke,C.
  TITLE     OLIGOCISTRONIC EXPRESSION SYSTEM FOR THE PRODUCTION OF HETEROMERIC
            PROTEINS
  JOURNAL   Patent: WO 9811241-A 1 19-MAR-1998;
            VON HOEGEN ILKA (BE)
FEATURES             Location/Qualifiers
     source          1. .8298
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /cell_line="BHK-21"
                     /cell_type="FIBROBLAST"
     promoter        1. .904
     intron          905. .976
     CDS             977. .>1018
                     /codon_start=1
                     /transl_table=11
                     /product="LEADER SEQUENCE (PART)"
                     /protein_id="CAB42566.1"
                     /db_xref="GI:4774642"
                     /translation="MGWSCIILFLVATA"
     intron          1019. .1106
                     /function="5'UTR POLIOVIRUS"
     CDS             <1107. .>1433
                     /function="FRS, CDRS"
                     /codon_start=1
                     /transl_table=11
                     /product="LIGHT CHAIN HMAB425, VARIABLE REGION, PLUS
                     LEADER(REST)"
                     /protein_id="CAB42567.1"
                     /db_xref="GI:4774643"
                     /translation="VHSDIQMTQSPSSLSASVGDRVTITCSASSSVTYMYWYQQKPGK
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:16 ; Search time 1487.61 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: US-09-980-548-1
Perfect score: 270
Sequence: 1 gctagcttaagtaacgccat.....cgccctcagcagtttctaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	270	100.0	270	6	AX040909	AX040909 Sequence
2	270	100.0	3123	12	PMBC2TD	X77750 Plasmid pMB
3	270	100.0	8298	6	A70359	A70359 Sequence 1
4	255.6	94.7	646	14	REMSV4	V01183 Moloney mou
5	255.6	94.7	903	14	MLMTRS01	J02268 Moloney mur
6	255.6	94.7	903	14	REMSV3	V01182 Left long t
c 7	254.6	94.3	3263	6	AX002278	AX002278 Sequence
c 8	254.6	94.3	3263	6	E28258	E28258 Purificatio
c 9	254.6	94.3	12516	12	XXU13843	U13843 pBPV clonin
10	246	91.1	6575	6	AX449148	AX449148 Sequence
11	223	82.6	2854	14	MSYMOS	K01683 Myeloprolif
c 12	209	77.4	544	10	S66424	S66424 c-myc [5' L
13	208.2	77.1	4148	6	AX166278	AX166278 Sequence
14	208.2	77.1	9318	6	AR028672	AR028672 Sequence
15	206.4	76.4	733	14	MLM3LTR	M24204 Moloney mur
16	206	76.3	5893	14	REAMLV	V01541 Abelson mur
17	206	76.3	5894	14	AF033812	AF033812 Abelson m
18	206	76.3	5894	14	MLAPRO	J02009 Abelson mur
19	202.8	75.1	508	6	AX107878	AX107878 Sequence
20	202.8	75.1	594	14	MMSLTR1	J02258 Moloney mur
21	202.8	75.1	635	6	AR016499	AR016499 Sequence
22	202.8	75.1	635	6	AR096882	AR096882 Sequence
23	202.8	75.1	636	6	E05956	E05956 Promoter/en
24	202.8	75.1	702	6	I77226	I77226 Sequence 16,
c 25	202.8	75.1	1200	10	RATDLTR	M19096 Rat DNA wit
26	202.8	75.1	1860	10	AY133244S3	AY133246 Rattus no
27	202.8	75.1	3097	6	AX107931	AX107931 Sequence
28	202.8	75.1	4207	6	AX359930	AX359930 Sequence
29	202.8	75.1	4207	6	AX382144	AX382144 Sequence
30	202.8	75.1	4210	6	AX359931	AX359931 Sequence
31	202.8	75.1	4210	6	AX382145	AX382145 Sequence
32	202.8	75.1	4273	6	AR112508	AR112508 Sequence
33	202.8	75.1	4273	6	AR201984	AR201984 Sequence
34	202.8	75.1	4273	12	PLTRPOLY	X60280 Vector plas
35	202.8	75.1	4661	6	AX359936	AX359936 Sequence
36	202.8	75.1	4661	6	AX382150	AX382150 Sequence
37	202.8	75.1	4776	6	A95117	A95117 Sequence 1
38	202.8	75.1	4776	6	A95152	A95152 Sequence 1
39	202.8	75.1	4922	6	AX146812	AX146812 Sequence
40	202.8	75.1	5109	6	I56770	I56770 Sequence 1
41	202.8	75.1	5117	12	AF132210	AF132210 Cloning v
42	202.8	75.1	5130	6	AX359935	AX359935 Sequence
43	202.8	75.1	5130	6	AX382149	AX382149 Sequence
44	202.8	75.1	5176	6	I56771	I56771 Sequence 2
45	202.8	75.1	5177	6	AX018983	AX018983 Sequence

ALIGNMENTS

RESULT 1
AX040909
LOCUS AX040909 270 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 1 from Patent WO0065074.
ACCESSION AX040909
VERSION AX040909.1 GI:11340531
KEYWORDS .
SOURCE Myeloproliferative sarcoma virus.
ORGANISM Myeloproliferative sarcoma virus
Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.
REFERENCE 1 (bases 1 to 270)
AUTHORS Mueller,P., Geserick,C., Schroeder,K. and Hauser,H.
TITLE Promoter-transactivator system for inducible high-level mammalian gene expression with the option of cell growth control